ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

PROCESS MINING FOR ANALYSIS OF INDOOR CUSTOMER BEHAVIORS

Ph.D. THESIS

Onur DOĞAN

Department of Industrial Engineering

Industrial Engineering Programme

MAY 2019

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

PROCESS MINING FOR ANALYSIS OF INDOOR CUSTOMER BEHAVIORS

Ph.D. THESIS

Onur DOĞAN
(507132115)

Department of Industrial Engineering

Industrial Engineering Programme

Thesis Advisor: Assoc. Prof. Başar ÖZTAYŞİ

MAY 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

SÜREÇ MADENCİLİĞİ İLE İÇ MEKAN KULLANICI DAVRANIŞLARI ANALİZİ

DOKTORA TEZİ

Onur DOĞAN
(507132115)

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Tez Danışmanı: Assoc. Prof. Başar ÖZTAYŞİ

MAYIS 2019

Onur DOĞAN, a Ph.D. student of ITU Graduate School of Science Engineering and Technology 507132115 successfully defended the thesis entitled "PROCESS MINING FOR ANALYSIS OF INDOOR CUSTOMER BEHAVIORS", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Başar ÖZTAYŞİ** Istanbul Technical University	....
Jury Members :	**Prof. Dr. Cengiz KAHRAMAN** Istanbul Technical University	....
	Prof. Dr. Selçuk ÇEBİ Yıldız Technical University	....
	Assoc. Prof. Sezi ÇEVİK ONAR Istanbul Technical University	....
	Assoc. Prof. A. Çağrı TOLGA Galatasaray University	....

Date of Submission : **06 May 2019**
Date of Defense : **20 May 2019**

To my spouse Deniz and daughter Defne,

FOREWORD

Interest in data science is growing day by day. Many researchers, scientists, and experts regard data science as the specialty of the future. In the 1970s, such as the emergence of computer science, recently, undergraduate and graduate programs and research centers working on data science have become widespread. However, the need for data science has gained more importance with a large amount of data brought by the industry 4.0 age. The data enable organizations to make the right decision and thus stay alive in a competitive environment. Results that cannot be seen at first glance can be revealed by data sciences. However, it is not enough to store and analyze the data. A data scientist should be able to master business processes and manage the process using the right questions and the right data. Process mining establishes the link between traditional process management and data-centered analysis techniques. The event data used in process mining enables the identification of problems in real processes rather than the structured processes.

Process mining makes real decisions as it uses the data recorded with data collection technologies in the system. Business intelligence tools focus on dashboards, tables or reporting rather than on clearly defined business processes. Data mining techniques are data centered. The results do not help understand the process. Therefore, only data mining techniques are insufficient to understand customer behaviors in indoor locations. Process mining method has been preferred because of the purpose of finding the routes followed by the customers, the time spent in the places they visited, the bottlenecks and the crowds.

Data collection type is critical for process mining applications to analyze processes. For online businesses, collecting data is easier than physical (brick-and-mortar) businesses. Online businesses can produce personalized recommendations to help customers come back; in other words, understand customer behavior. Each click or movement of the cursor is saved, and businesses send receipts, codes, advertisements or other material based on these data. In physical stores, many alternative technologies have been developed to collect data to solve this problem. Radio frequency identification (RFID), wireless connection (WiFi - wireless fidelity), Bluetooth are the most preferred ones. Different technologies have been developed to collect data in the interior, but there is no consensus among them. Each method has its advantages and disadvantages.

In this study, we first decided the most suitable data collection technology with a fuzzy decision-making method. Bluetooth was selected and established in a shopping mall in Istanbul. Then we collected data from customers' smartphones to analyze their indoor paths with process mining. We evaluated male and female customer behaviors from the paths discovered by PALIA algorithm, which is a process discovery algorithm in process mining. The results showed that male and female customer have different behavior in the shopping mall.

I would like to thank everyone who contributed to the studies that resulted in long efforts and endeavors. First of all, I would like to thank Assoc. Prof. Başar

ÖZTAYŞİ, who guided and helped me with her knowledge and experience in shaping and publishing my studies. Additionally, Prof. Cengiz KAHRAMAN and Prof. Selçuk ÇEBİ was supported my studies by their meaningful opinions.

I want to express my special thanks to ITACA SABIEN Research Group in Spain, especially Dr. Carlos FERNANDEZ-LLATAS, who is the scientific coordinator and developer of PALIA Suite, for his unlimited interest and support that he showed me in the solution of my process mining application.

Also, I would like to express my gratitude to The Scientific and Technological Research Council of Turkey (TUBITAK) for the financial support that has given under the 2214-A International Ph.D. Research Scholarship Program in the realization and publication of the studies.

I am also thankful to the other members of my wife Deniz DOĞAN for their motivation during this challenging Ph.D. process.

May 2019

Onur DOĞAN
(Industrial Engineer)

TABLE OF CONTENTS

Page

ABBREVIATIONS

AHP	: Analytic Hierarchy Process
ARM	: Association Rule Mining
ATM	: Automatic Transfer Machine
BPMN	: Business Process Management Notations
BLE	: Bluetooth Low Energy
CBA	: Customer Behavior Analysis
CRM	: Customer Relationship Management
EPC	: Event-Driven Process Chain
ERP	: HEnterprise Resource Planning
FPGrowth	: Frequent Pattern Growth
HFLTS	: Hesitant Fuzzy Linguistic Term Sets
HFS	: Hesitant Fuzzy Sets
IBAT	: In-store Behavioral Analytics Technology
ILS	: Indoor Location System
IR	: Infrared
MAC	: Media Access Control
MCDM	: Multi-criteria Decision Making
OWA	: Ordered Weighted Averaging
PALIA	: Parallel Activity Log Inference Algorithm
PM	: Process Mining
RFID	: Radio-frequency identification
SCM	: Supply Chain Management
SDK	: Software Development Kit
TPA	: Timed Parallel Automatons
TrFN	: Trapezoidal Fuzzy Number
WiFi	: Wireless Fidelity

LIST OF TABLES

LIST OF FIGURES

Page

PROCESS MINING FOR ANALYSIS OF INDOOR CUSTOMER BEHAVIORS

SUMMARY

The most important issue for maintaining the existence of an enterprise is to increase the profit by creating customer loyalty. The main concern for this is to understand customers. Customers expect to see personalized advice from companies like Netflix, Amazon, and Spotify in stores. In a recent study by Infosys, a retailer's offer of customer-oriented and personalized offers indicated that 78% of consumers would be repeat customers. In another similar study, Microsoft says that 73% of customers prefer to shop with brands that offer a personalized purchase experience. By 2020, the customer experience, price, and product/service will be passed as a brand separator. This demonstrates the importance of understanding customers.

For online businesses, it is easier to produce personalized recommendations to help customers come back (understand customer behavior). Each click or movement of the cursor is saved, and businesses send receipts, codes, advertisements or other material based on these data. Physical stores (brick-and-mortar) should also provide customers with an experience that matches the specific service they receive online and must relate to digital channels seamlessly. According to RetailNext, 84% of customers believe that retailers should make more efforts to integrate online and offline channels.

The integration problem of online and offline channels can be solved by behavior analytics. Behavioral analyses investigate to understand the needs of patients and to provide the best solutions to these needs. As in online shopping, it is also possible to present personalized patient recommendations in shopping malls by behavioral analysis. The behavioral analysis may provide information about the stores in the shopping mall such as the time spent in the stores, the next visit of customers, and how many people have looked at the screen without entering into the store. For example, personalized offers such as sending an advertisement to the customer by followed path, reviewing the past visits and giving an individualized discount increase customer loyalty. However, in all physical stores, including a shopping mall, collecting enough data to recognize customers is an important problem.

There are many alternative technologies have been developed to collect data to solve this problem in physical stores. Radio frequency identification (RFID), wireless connection (WiFi - wireless fidelity), Bluetooth are the most preferred ones. Apart from these, classical methods (observation, questionnaire, interview) and wireless communication technology (infrared, ZigBee and computer vision technologies such as camera and motion sensor) are the technology alternatives that can be used indoors to collect data. Fuzzy multi-criteria decision-making method is applied for the selection of data collection technology in this study. As a result, real data will be obtained without the human factor (without manual data entry) and the actual results will be analyzed by analyzing these data.

Analyzing the data obtained with the selected data collection technology is very important in identifying patient behavior in the best way. It is another important problem to follow the paths within the shopping mall to understand the customers. When the number of followed customers and the number of indoor stores increase, the hardness of the problem gets bigger. Therefore, the selection of the analysis method is also another critical step. Implementing basic and advanced statistical rules in processes in indoor systems is not sufficient to give an overview of the processes. In order to use the advantage of the data collected, new technologies are needed to facilitate the process characteristics for people to understand and to give information about the whole process. Process mining is a new way of using event logs to generate models that people can read. When we look at the application areas of process mining, it is seen that it is mostly used for the process flow improvements in many businesses. With the shopping mall data collected by the indoor systems, the process of the followed paths by the customers and the time spent in the stores where the customer visits are considered as one of the original aspects of this research. Thus, automatically collected data will be automatically used for customer behavior analysis. In some analysis methods, an expert is needed to interpret the results. One of the benefits of process mining is that the results are easily understandable by non-experts. Process mining makes real decisions as it uses the data recorded with data collection technologies in the system. Business intelligence tools focus on dashboards, tables or reporting rather than on clearly defined business processes. Data mining techniques are data centered. The results do not help understand the process. Therefore, only data mining techniques are insufficient to understand customer behaviors in indoor locations. Process mining method has been preferred because of the purpose of finding the followed paths by the customers, the time spent in the places they visited, the bottlenecks and the crowds.

In this study, we first decided the most suitable data collection technology with a fuzzy decision-making method. Bluetooth was selected and established in a shopping mall in Istanbul. Then we collected data from customers' smartphones to analyze their indoor paths with process mining. We evaluated male and female customer behaviors from the paths discovered by PALIA algorithm, which is a process discovery algorithm in process mining. The results showed that male and female customer have different behavior in the shopping mall.

SÜREÇ MADENCİLİĞİ İLE İÇ MEKAN KULLANICI DAVRANIŞLARI ANALİZİ

ÖZET

Bir işletmenin varlığını sürdürmesi için en önemli konu müşteri bağlılığını oluşturarak karı artırmaktır. Bunun için temel mesele müşterileri anlamaktır. Müşteriler Netflix, Amazon ve Spotify gibi şirketlerden edindikleri kişiselleştirilmiş önerileri mağazalarda da görmeyi beklemektedirler. Yakın zamanda Infosys şirketi tarafından yapılan araştırmada, bir perakendecinin müşteri hedefli ve kişiselleştirilmiş teklifler sunmasının, tüketicilerin %78'inin tekrar müşteri olacağını bildirdi. Benzer bir başka araştırmada da Microsoft, müşterilerin %73'ünün, kişiselleştirilmiş bir satın alma deneyimi sunan markalarla alışveriş yapmayı tercih ettiğini söylüyor. 2020'ye gelindiğinde, bir marka ayırıcı olarak müşteri deneyimi, fiyat ve ürünün/hizmetin kendisini geçecektir. Bu durum da müşterileri anlamanın önemini açıkça ortaya koymaktadır.

Çevrim içi satış yapan işletmeler için müşterilerin geri gelmesini sağlayacak kişiselleştirilmiş öneriler üretmek (müşteri davranışlarını anlamak) daha kolaydır. Her tıklama veya imlecin hareketi kaydedilir ve işletmeler bu verilere dayanılarak fiş, kod, reklam veya başka materyal gönderir. Fiziki mağazalar (brick-and-mortar) da, müşterilere çevrim içi aldıkları özel hizmetle eşleşen bir deneyim sunmalı ve dijital kanallarla sorunsuz bir şekilde ilişki kurmalıdır. RetailNext'e göre, müşterilerin %84'ü perakendecilerin çevrimiçi ve çevrimdışı kanalları entegre etmek için daha fazla çaba göstermesi gerektiğine inanıyor.

Çevrim içi ve çevrim dışı kanalları entegrasyon sorunu davranış analizi ile çözülebilir. Davranış analizleri, hastaların ihtiyaçlarını anlayıp, bu ihtiyaçlara en iyi çözümler sunmak için yapılan analizlerdir. Çevrim içi alışverişte olduğu gibi alışveriş merkezinde de kişiselleştirilmiş müşteri önerileri sunmak davranış analizi ile mümkün hale gelir. Davranış analizi, müşterilerin AVM içinde uğradığı yerler, bu yerlerde harcadıkları süre, sonraki ziyaret yeri ve hatta kaç kişinin vitrinden bakıp içeri girmeden gittiğine dair bilgiler üretilebilir. Örneğin, AVM içinde 60 dakikadan fazla bekleyen müşterilerin acıktığı düşünülürse, geçmiş rotaları gözden geçirilerek müşteriye özel yemek indirimleri vermek gibi kişiselleştirilmiş teklifler müşteri bağlılığını oluşturmaktadır. Ancak alışveriş merkezleri dahil genel olarak tüm fiziki mağazalarda, müşterileri tanıyabilecek kadar veriyi toplamak önemli bir problemdir.

Alışveriş merkezlerinde veri toplamak için kullanılabilecek birçok teknoloji alternatifi vardır. RFID, WiFi, Bluetooth bunlardan en çok tercih edilenleridir. Bunların dışında gözlem, anket, mülakat gibi klasik yöntemler, kızıl ötesi, zigbee gibi kablosuz iletişim teknolojisi ve kamera, hareket sensoru gibi bilgisayarla görme (computer vision) teknolojileri de veri toplamak için iç mekanlarda kullanılabilecek teknoloji alternatifleridir. Bu araştırmada RFID, WiFi veya Bluetooth teknolojilerinden biri ile veri toplanacaktır. Veri toplama teknolojisinin seçimi için bulanık çok kriterli karar verme (fuzzy multiple criteria decision making) yöntemi uygulanmıştır.

Veri toplama teknoloji seçim problemi, dört kriter ve sekiz alt kriter içeren bir çok kriterli karar verme problem olarak modellenmiştir. Geniş bir literatür ve uzman görüşmesinden sonra, bu problemin alt kriterleri Maliyet, Zaman, Doğruluk, Önyargı, Miktar, İntrospektif, Sadelik ve Analiz Kolaylığı olarak belirlendi. Bu kriterler arasında en önemlisi Doğruluk, Önyargı ve Maliyettir.

Bu çalışma, Bluetooth teknolojisinin yapılan uzman değerlendirmelerine göre diğer veri toplama teknolojileri arasında en iyi alternatif olduğunu ortaya koymuştur. Wi-Fi, Bluetooth'a yakın bir puan alan en iyi ikinci alternatiftir. Bu yakın puan, teknolojideki yeni bir gelişmenin iki alternatifin sırasını değiştirebileceğini ortaya koymaktadır. Aktif olarak tedarik zinciri, lojistik ve depo uygulamaları için kullanılan RFID teknolojisi üçüncü sırada yer almaktadır. RFID'in üçüncü sırayı almasının ana nedenleri ön yükleme için etiket yatırımı gerektirmesi ve ayrıca ziyaretçilerin bu etiketleri taşımasına direnmesi olabilir. Kamera sistemleri, başlangıçtaki ilk yatırım ve yüksek işletme maliyetleri nedeniyle sıralamada son sırada yer almaktadır. Sonuç olarak, Bluetooth ile insan faktörü olmadan (elle veri girişi yapılmadan) gerçek veri elde edilmiş ve bu veriler süreç madenciliği için kullanılmıştır.

Seçilen veri toplama teknolojisiyle elde edilen verilerin analiz edilmesi, müşteri davranışını en iyi şekilde tanımlamada çok önemlidir. Müşterileri anlamak için alışveriş merkezi içindeki rotalarını takip etmek (hangi mağazaya uğradı, hangi mağazada ne kadar süre harcadı, sonraki ziyareti nereye oldu) önemli başka bir problemdir. Kapalı alan büyüdükçe veya takip edilecek hasta sayısı arttıkça problem daha büyük hale gelmektedir. Bu yüzden verilerin toplanması kadar analiz yönteminin seçimi de kritik bir adımdır. İç mekan sistemlerindeki süreçlerde temel ve ileri istatistik kurallarını uygulamak, süreçlerin anlaşılması konusunda genel bir bilgi vermek için yeterli olmaz. Toplanan verilerin avantajını kullanabilmek için, süreç karakteristiğini insanların anlaması için kolaylaştıracak, sürecin tamamı hakkında bilgi verecek yeni teknolojilere ihtiyaç vardır. Süreç madenciliği, insanların okuyabileceği modeller çıkarmak için süreçlerdeki kayıtları (event logs) kullanan yeni bir yöntemdir. Süreç madenciliğinin bir çok uygulama alanlarında, süreç akışlarının çıkarılması ve iyileştirmelerin yapılması için kullanıldığı görülmektedir. İç mekan sistemleriyle toplanan veriler sayesinde, müşterinin takip ettiği rotanın ve uğradığı yerlerde geçirdiği süre bilgisinin süreç madenciliği ile ele alınması bu araştırmanın orijinal yönlerinden biridir. Böylece otomatik olarak toplanan veriler, otomatik olarak müşteri davranış analizi için kullanılır hale gelecektir. Bazı analiz yöntemlerinde sonuçların yorumlanması için yine bir uzmana gerek varken, süreç madenciliğinin faydalarından biri de, elde edilen sonuçların uzman olmayan kişiler tarafından da kolaylıkla anlaşılabilir görsel çıktılar sunmasıdır.

Süreç madenciliği çalışmasında, bir alışveriş merkezinde erkek ve kadın müşteri yollarını keşfetmek için bir süreç madenciliği uygulaması sunulmaktadır. Bu metodoloji, Aralık 2017 - Şubat 2018 arasındaki gerçek üç aylık müşteri verileri kullanılarak test edilmiştir. Gerçek veriler, iBeacon cihazlarıyla Bluetooth teknolojisi kullanılarak toplanmıştır. Bluetooth, tarafsız ve etkilenmeyen gözlemler sağlayan uygun maliyetli bir izleme teknolojisidir. Nadir ve eksik bir olay günlüğü ile başa çıkabilecek bir süreç madenciliği algoritması olan PALIA Suite geliştirilmiştir. PALIA Suite seyrek davranışı, faaliyetlerin sıklıkları ve durum değişiklikleriyle ilgili istatistiksel bilgilerle değerlendirebilir. Çalışmanın sonuçlarına göre, iç mekan sistemleri ile proses madenciliğinin uygulanması anlaşılabilir bir görüş sağlamaktadır.

Temel istatistikler ve süreç madenciliği ile ilgili sonuçları değerlendirdik. Temel istatistikler bazı veri merkezli sonuçlar verirken, süreç madenciliği süreç yönelimlidir. 165 erkek, 477 bayan müşteri olan 642 müşteri rotasını araştırdık. Bazı müşteriler alışveriş merkezini bir kereden fazla ziyaret ettiğinden, 1293 müşteri rotası çıkardık. Bu çalışmada ziyaret edilen mağaza gruplarının sayısı 2749 idi. Aralık ayında alışveriş merkezini daha fazla müşteri ziyaret etti ancak Ocak ayında daha fazla mağaza ziyareti gerçekleşti. Şubat, müşteri sayısı ve ziyaret edilen mağaza sayısı bakımından en zayıf aydı. 9 Aralık ve 15 Şubat ekstremum tarihleriydi. 25 müşteri 9 Aralık'ta 59 mağaza grubunu ziyaret ederken, 7 müşteri 15 Şubat'ta 10 mağaza grubunu ziyaret etti.

Aralık ayında yeme-içme ve giyim en popüler mağaza gruplarıydı. Yeme-iöme ve giyim mağazaları en çok ziyaret edilenler ve ziyaret süreleri en fazla olan gruplardır. Giyimden sonra yeme-içme mağazalarını ve yeme-içmeden sonra giyim mağazalarını ziyaret eden müşterilerin sayısı sırayla erkek müşteri ziyaretlerinde 16 ve 14, kadın müşteri ziyaretlerinde ise 35 ve 47 idi. Erkek müşteriler, giyim ile yeme-içme ve giyim ile süpermarketi arasında bir döngü yaşadılar. Öte yandan, kadın müşterilerin yalnızca giyim ile yeme-içme döngüsü var. Bu durum, alışverişte kadın müşterilerin erkek müşteriden daha kararlı olduğu anlamına gelebilir. Ayrıca erkek müşteriler daha az sayıda mağaza grubunu ziyaret ettiğinden, alışverişte daha az seçici olduğu sonucuna varabiliriz. Aksesuar kadın ve erkek müşterilerin ziyaret ettiği ortak bir mağaza grubu olmasına rağmen, kadın müşteriler buralarda daha az zaman harcadılar. Ayrıca, eğlence ve kişisel bakımın cinsiyetler arasındaki davranışsal farkla ilgili önemi vardı. Erkek müşteriler, en çok izlenen yollarda herhangi bir eğlence veya kişisel bakım mağazasını ziyaret etmedi.

Ocak ayında yeme-içme ve giyim, Aralık ayında olduğu gibi erkek müşteriler için en popüler mağaza gruplarıydı. Diğer taraftan, kadın müşteriler için eğlence mağazaları ziyaretinin süresi artmıştır. Bunun nedeni, merakla beklenen bazı filmlerin vizyona girmesi olabilir. Kadın müşteriler çoğunlukla giyim mağazalarını ziyaret etse de, eğlence mağazaları için daha fazla zaman harcadılar. Birçok durumda, eğlence mağazalarını ziyaret eden müşteriler atıştırmak veya karınlarını doyurmak için bir yeme-içme mağazasını ziyaret etti. Erkek müşteri rotalarındaki döngüler Aralık'ta olduğu gibi hala kadın müşteri yollarından daha yüksekti. Aralık ve Ocak ayları arasında karşılaştığımızda, erkek müşteriler yine yeme-içme ve giyim mağazaları arasında döngüye sahipti. Ayrıca, Ocak ayında elektronik, anne ve bebek mağazaları, erkek müşteri ziyaretlerinde daha uzun bir süreye sahip oldu. Ev eşyaları mağazaları, kadın müşteriler için Aralık ayındakinden farkıydı. Ev eşyaları mağazala ziyareti sayısı neredeyse aynıydı, ancak ziyaret süresi Aralık ayındaki süreden dört kat daha fazlaydı.

Şubat ayında erkek müşteriler giyim yerine elektronik mağaza gruplarını ziyaret etmeyi tercih ettiler. Yeme-içme hala en çok ziyaret edilen yerlerden biriydi. Erkek rotalarındaki ziyaret edilen mağazalar ile harcanan süre arasında ters orantı olması ilginçtir. Örneğin, erkek müşteriler yeme-içme ve giyim mağazalarını daha fazla sayıda ziyaret ettiği halde buralarda daha az süre harcadılar. Dieğr taraftan elektronik mağazaları daha az ziyaret edip daha fazla zaman harcadılar. Diğer iki ayın aksine, kadın müşteriler Şubat ayında daha fazla döngülere sahipti. Kadın müşteriler bu ay yiyecek içecek yerine aksesuar mağazası gruplarını ziyaret etti. Bu durum, Sevgililer Günü nedeniyle olabilir. Ancak, erkek müşteriler o özel günde hediye almakla ilgilenmediler.

Genel olarak, yeme-içme mağazalarında daha fazla zaman harcayan müşterilerin giysi için daha az zaman harcaması ilginçtir. Erkek ve kadın müşterilerin benzer rotaları olmasına rağmen, harcanan süreler iki grup için farklıydı. Yeme-içme, erkek ve bayan müşteriler için çoğunlukla en yüksek süreye sahipti. Ancak, eğlence ve aksesuar gibi bazı mağaza grupları, bazı durumlarda davranışsal değişiklikler için kritik nokta haline geldi.

1. INTRODUCTION

Data science is a discipline involving several disciplines such as databases, statistics, data mining, machine learning, artificial intelligence. Current approaches in these disciplines aim to reveal valuable information for organizations and the community by using abundant and meaningless data. With the start of the Industrial 4.0 era, new challenges emerged in terms of both the size of the data and the questions to be answered. Now, the amount of data collected by the tools connected to the Internet has increased considerably. Achieving meaningful results from these data has become more and more important in a challenging competitive environment.

Society has recently been rapidly shifting from analog to digital [1]. Moving to a digital environment has been a change that has greatly affected business life [2]. People and companies are now connected to the Internet at any moment. According to a report by Turner and colleagues [3], 134 billion vehicles are connected to the internet. These reasons can be collected anytime and anywhere, about anything data. The collected data are increasing each year with the exponential growth rate. Developments in mobile devices, digital sensors, communication, information processing, and storage have facilitated data collection [4]. At the 2012 World Economic Forum in Davos, it has been reported that large data has become a similar strategic economic resource in terms of foreign exchange and gold [5]. Walmart can be given as an example of magic data generation. Walmart processes and takes over 1 million customers to the database; this value contains more than 2.5 petabytes of data per hour [4]. One of the most important problems of today's firms is to extract meaningful data from information systems and transform it into information [6].

The increase in the importance of information systems is not only the growth of the amount of data but also the inclusion of today's digital and physical environments. For example, the price we pay for a flight ticket we buy from the internet is also digital. The ticket is an electronic document consisting of numbers only. Another example is the RFID (Radio Frequency Identification) tag of a product sold to the customer, and it is possible to reduce the number of products in stock by ERP (Enterprise Resource Planning) software. In fact, the product sold was physically in stock but sold digitally to the customer [6].

There is a need for scientific approaches and methods for analyzing and monitoring human behavior and for personalizing processes. In this targeted research, the main scientific aim is to present human behavior in a specific shopping mall and to present an individualized behavioral model to understand the customers. Actual data must be collected for accurate behavior analysis. A fuzzy multi-criteria decision-making approach will be developed for the selection of the previously mentioned data collection technologies. With the selected data collection technology, data mining techniques will be collected and analyzed.

Before process mining, we applied association rule mining to investigate stores frequently visited together. The stores are divided into groups according to the products they sell or the services they provide because store group data are more meaningful and easy to interpret regarding the aim of the study.

In the proposed research, current technologies and algorithms will be used for human behavior analysis to be combined with process mining. Customer data to be used as input to process mining will be collected with the help of the technology at the end of the fuzzy decision-making method. Algorithms to be applied as prototypes especially in the field of health applications.

In the retail domain, technology selection for data collection and process mining for data analysis are combined for the first time in human behavior analysis. In the research, some problem topics are discussed in different stages of data collection, analysis and implementation. Selection of data collection technology is the first topic. We try to find answers to the question: a) what are the advantages and disadvantages of determined technologies in data collection? b) What are the evaluation criteria to be considered in the evaluation of technologies? How should these criteria be determined? c) Which multi-criteria decision-making method will be applied?

The second main topic of the study is the application of process mining method be considering the questions: a) Is the status of the collected data suitable for the method? Are there data that need to be cleared, such as noise, outliers? b) Which of the process mining algorithms will be applied? c) Does one of the available softwares yield results? If necessary, will be developed a new tool specific to the problem?

In the process mining application, in order to prepare and validate the developed model for all stores in the shopping mall, individual behaviors of some randomly selected customers are examined. Here are some significant points to consider: a) The followed path by each customer from any store to any store. It is not necessary to start and finish at the entrance because the paths do not represent exactly a business process. b) The spent time every store customers visit c) Personal information of the customer if it is possible to create from the data.

The general aim of the study is to select the data collection technology to be used in the retail sector by fuzzy multi-criteria decision-making method and to develop a current customer behavior analysis method which performs the analysis of the collected data with process mining.

1.1 Purpose of Thesis

The research area is limited to indoor locations. More specifically, it is applied to a shopping mall in Istanbul. Analysis of customer behaviors is necessary for use in more effective ways of resources and marketing information to provide more effective responses to their needs.

The main purpose of this study is to develop a behavioral model using process mining. The data required for the development of the model is based on monitoring customer paths. Proper customer behaviors require to follow their movements within the indoor locations. Therefore, technology must be selected to track the customers as well as to collect spatio-temporal data. Alternative data collection and tracking technologies are determined, and their evaluations are made in terms of different criteria. Previously the selection of the appropriate technology for data collection is not made with a scientific study. In this study, technology selection is made by applying a fuzzy multi-criteria decision-making method. Thus, it is aimed to collect the necessary data for customer behaviors with the most appropriate technology.

One of the study goals is to implement an effective method by using technologies, databases and defined criteria in process mining. This method enables very usefully visualized results to see realized situations rather than what is desired in customer behavior. Visualization of the thousands of movements of customers within the shopping mall, generalizing these paths, realizing the crowds in the stores or making the results understandable that everyone can understand can be obtained by process mining.

In addition to the aims of the study, the research is a good opportunity to introduce process mining which is not known in Turkey but is frequently used in hospitals, municipalities, factories, and offices around the world.

In summary, the study aims to select the most appropriate technology to collect indoor customer data in order to analyze the customer behaviors from the paths using process mining method. Additionally, we analyze the visiting data of a shopping mall collected via Beacon devices, which have Bluetooth technology, the popularity of the stores and the most frequent routes. The results reveal that the locations on frequent routes are not always popular. In other terms, frequent routes are different from popular locations.

1.2 Importance of the Research and Originality

The retail sector is under pressure from employees due to insufficient vocational training, inadequate human resources, the number of people in different situations, the high number of operations and the lack of good management and customer-oriented service [7]. With personalized behavior analysis, the store managers will recognize easier customers and thus provide customer-oriented services. There are several applications available for this situation. One of them is to follow up with the latest technologies within the store. No preliminary studies have been conducted for the selection of the tracking technology. The application of hesitant fuzzy multi-criteria decision-making method to select the most appropriate tracking technology is a processor study for those who will focus on a similar investigation. Research is methodologically important in this respect. On the other hand, it is a situation that store managers can know their customers with behavior analysis and increase the operating profit by providing customer loyalty.

The research has two potential target groups: store managers and customers. Store managers work to better understand their customers and address their needs. On the other hand, the dependence of the customer who feels special increases. This situation has material and moral (brand reputation, reputation, customer loyalty, etc.) returns to company management by establishing a positive relationship.

Turkey's retail market, after Russia and Poland, is one of the most vibrant and dynamic countries in Europe [8]. Moreover, Turkey is a European country with a high propensity to consume [9]. For these reasons, it is critical to understand the needs of customers for retail companies that want to get the highest share in meeting the current customer demand. Exchange rate fluctuations and retailers with the reasons for the decline in consumer confidence and high profitability rather than sales growth and are focused on long-term investments. From this point of view, the concept of understanding the customer especially in ours has emerged as a strategy for the retail sector facing competition [10]. Understanding customers' shopping behaviors and profiles are becoming more and more important. As the world consumer profile in Turkey, which determines consumers' viewing habits and behaviors and strategies based on the results of these investigations retailers and manufacturers, they pass an advantage against competitors.

Process mining (PM) is a technique that aims to extract understandable information via process models generated automatically by different discovery algorithms from event logs using process management and business analytics [11, 12]. Most of the pathway analyses use graphical representation to show the pathways. The created pathways are intuitive without any detailed study [13]. However, process mining explains the underlying information of the pathway. Therefore, process mining is selected to discover pathways of indoor customers to understand the behavior of customers visiting the shopping mall. In the process mining study, behavioral analysis refers to understanding customer needs and proposes the best solutions to meet these needs. The study does not aim to investigate behavioral science related to retail.

2. IN-STORE BEHAVIORAL ANALYTICS TECHNOLOGY SELECTION USING FUZZY DECISION MAKING[1]

Advances in technology have enabled collecting, storing and processing a large amount of data. Automated collection of human behavior is one of the recent developments in data collection. Companies can analyze the behavior data to get insight into the interests and desires of individuals [14]. The technology can be used to detect movements of employees to understand common routes and optimize facility location in production systems; it can also be used to monitor potential customers in stores for marketing purposes. The technology is very important for customers to understand and communicate with their customers. Nowadays, customers expect to see personalized proposals from both production and service companies especially stores such as companies like Netflix, Amazon, and Spotify [15, 16]. In a recent study by the Infosys Company, it was reported that 78% of consumers would be customers again if a retailer offers customer-targeted and personalized offers [17]. In another similar study, Microsoft states that 73% of customers prefer to shop with brands that offer a personalized buying experience [18]. By 2020, customer experience is expected to pass price and product as a brand separator [19]. These findings underline the importance of understanding customers.

It is easier to generate personalized recommendations that allow customers to return to Internet-based businesses. Each click or cursor movement is recorded, and the operators send the chip, code, advertisement or other material based on this data. According to RetailNext, 84% of customers believe that retailers need to do more to integrate their online and offline channels [20]. Accordingly, physical stores should provide customers with an experience that matches the specific service they receive from the Internet and should establish an unproblematic relationship with digital channels [21]. The problem of integrating online and offline channels has been solved by tracking customers in the store using different technologies [16, 21–23].

[1]This chapter is based on the paper Dogan, O., and Oztaysi, B. (2018). In-store behavioral analytics technology selection using fuzzy decision making. *Journal of Enterprise Information Management, 31*(4), 612–630.

The traditional method to understand customer behavior is to use questionnaires, interviews, and observations. With the development of technology, new data collection methods have emerged. The common feature of the new methods is that the quantity and accuracy of the collected data are high. RFID (Radio-frequency identification), WiFi (wireless fidelity), Bluetooth, infrared (IR), ZigBee [24, 25], and computer vision techniques such as camera [16, 26–28], motion sensor [29] and ultrasound are technology alternatives that can be used for collecting indoor location data. The RFID technology is a tool that collects and stores data by wireless communicating with tags on the object to be tracked [30]. WiFi is another indoor technology that determines the location of the user through the signals at the access points used [31]. Another alternative is Bluetooth technology, which allows the system to recognize and track users through the MAC (Media Access Control) address of their smartphones. A technology similar to Bluetooth is ZigBee which is widely preferred because it is low cost in applications that can be performed with small size data exchange. Computer vision technology, which collects data with image recording, uses advanced image processing algorithms with the complex structure to describe the object to be tracked [23]. The motion sensor technology analyzes high-cost multi-camera images that are sensitive to the indoor locations and monitor the target areas [32]. IR technology can also be used indoors in total darkness, as opposed to the motion sensor [33]. However, now that smartphones do not usually contain an infrared transmitter, establishing a closed space with this technology requires an expensive infrastructure [34]. Ultrasound technology is a technology that determines the location of an object by extracting sound waves that the human ear cannot hear [35]. An ultrasound sensor continuously sends sound waves. It calculates the distance of the object using the time that it takes to strike to the object and return to the sensor.

Selection of proper technology for indoor behavioral analytics is a multi-criteria decision making (MCDM) problem where there are multiple alternatives and criteria. Analytic Hierarchy Process (AHP) proposed by Saaty [36] is an MCDM method which is used for complex decision problems. In the original method, linguistic terms are represented as crisp numbers; however, later the method is extended by using fuzzy sets. Hesitant Fuzzy Sets (HFS) is an extension of ordinary fuzzy sets which can handle the hesitancy of the decision makers [37]. Hesitant Fuzzy Linguistic Term

Sets (HFLTS) are proposed based on HFSs to use linguistic terms with context-free grammar in decision-making problems [38]. Hesitant Fuzzy AHP method has been proposed by utilizing HFLTS in AHP method [39, 40]. In this study, Hesitant Fuzzy AHP method is modified to employ fuzzy trapezoidal sets while in the original studies triangular fuzzy numbers are used.

The rest of this paper is organized as follows. Section 2.1 presents a literature review on in-store behavioral analytics. Section 2.2 presents technology alternatives are introduced. In Section 2.3, evaluation criteria for the selection problem are given in detail. In Section 2.4, first, the preliminaries are presented, and then the steps of the methodology are explained. In Section 2.5, a real case which aims to select the most suitable technology for in-store analytics is given with sample calculations. Finally, the concluding remarks are given in Section 2.6.

2.1 Literature Review: In-store Behavioral Analytics

There are many different ways to collect the data needed to analyze customer behavior. The questionnaire method is the most commonly used data collection methods among other traditional methods such as observation, interview, and document review. Kirchberg and M. Tröndle [41] conducts interviews to investigate the similarities and differences between the experiences of visitors. Data are collected comparing visitors' expectations (motivations before the visit), experiences (social interactions during the visit) and contributions (what they learned after the visit). Operators are asked to directly observe the behavior of customers in the store if they need more details about the customer's behavior to set better the number of employees who need to be in their sales area [16].

Unlike traditional methods, it has become possible to combine much more data without touching through technology. One of the disadvantages of traditional methods is that the real time of activities is not entered. Thus, the analyzed data deviates from the actual results. Fernandez-Llatas et al. [42] analyze some processes in a hospital with process mining. For process mining, it is important to know the actual start and finish times of the activities. For this reason, they have collected and analyzed data with RFID technology. Yuanfeng et al. [43] have determined the locations of users in a closed area with signals sent to users' mobile phones from WiFi. Yoshimura et al. [44]

conduct a study to avoid the negative effects of the crowd in the museum over the experience of the visitors. In this study, the visitors are separated as to those who stayed short – medium – long according to collected data, and results obtained from the routes for the different group are discussed. Zigbee technology similarly collects data with Bluetooth. The difference is that ZigBee communicates with data exchange on a small scale. Huang and Chan [24] use the behavior data gathered by ZigBee technology to determine which object is closer than previously determined positions.

Using the data collected by the cameras are also used in the literature. Wu et al. [21] use the in-store camera recordings to understand the behavior of customers and the way they tracked. The ways customers follow are examined with a heat map. In another study, Park et al. [45] combine motion sensor technology with Bluetooth to compare the results of two technologies. A self-correction system to improve the accuracy of the tracking are developed for evaluation of the different results obtained from the two methods. Liu et al. [26] use infrared data to identify the characteristics of the target object with artificial neural networks in a completely dark environment. In another paper, Ran et al. [46] a precise position measurement system developed with ultrasound technology to guide blind users and help them to navigate independently and safely in indoor and outdoor that they know and do not know. A comprehensive research result on literature review is given in Table 2.1.

2.2 In-store Behavioral Analytics Technology (IBAT) Alternatives

There are many different ways to collect data to analyze customer behavior in indoor locations. Traditional data collection technologies such as questionnaires, interviews, observation, and document scanning are still preferred methods in many studies [41, 47–49]. Traditional methods are particularly useful in understanding the inner world of customers, especially where technological data collection methods are lacking. However, it is preferable that the practitioners are trained to understand the questions, how to approach the customers, how to evaluate the doubtful answers, how to finish the interview or the questionnaire [50]. This is also time-consuming compared to technological methods.

RFID is an automatic identification method based on collecting and storing data using devices called tags or transponders [51]. RFID systems use radio waves to switch from

Table 2.1 : Literature review of data collection for indoor locations.

Technology	General Information	Studies	Example Study
Traditional Methods	They express data collection with a questionnaire, interview, and observation.	[41, 47 - 50]	Visitor interviews were conducted to investigate the similarities and differences between visiting experiences in a museum [41].
RFID	RFID communicates with radio frequency readers through the tags of the object tracked.	[22, 27, 28, 30, 32, 42, 54, 56]	Actual results cannot be achieved due to the fact that data entries are not made immediately after the activity ends. To overcome this problem, RFID technology is used to gather data. Process mining has been applied to gathered data [42].
WiFi	WiFi-enabled devices regularly send messages with a unique identifier. A WiFi scanner retrieves these messages, which can then be processed. Having a unique identifier allows devices to be detected and tracked.	[31, 43, 58-61, 64]	A tracking and positioning system is proposed by collecting data with WiFi from mobile phones of users. The collected data are used for algorithms such as The Maximum Likelihood, Kalman Filter to provide flexibility, correctness, and efficiency in the system [43].
Bluetooth	Messages sent via Bluetooth contain the Bluetooth MAC address, which can be reused to identify the device. As a result, when enough scanners are placed, the position of the object can be accurately calculated.	[23, 44, 65-68]	The major problem in most museums is to control the crowd. The data collected via Bluetooth were grouped into groups such as short-medium-long stayed. Thus, some results were obtained to control the crowd in the museum [44].
Zigbee	It is a low-cost wireless communication technology based on the principle of minimum power consumption, which can be performed by data exchange in a small size.	[24, 25, 69]	The data collected from the object with Zigbee technology is a work that determines where the object is closest to the k-nearest algorithm [24].
Camera	It uses sophisticated image processing algorithms with the complex structure to describe the object to be tracked.	[16, 21, 26, 71-73]	The roads that customers follow are turned into temperature maps using the in-store camera recordings to understand the behavior of customers and the way they follow [21].
Motion sensor	Motion sensor technology analyzes high-cost, multi-camera images that are sensitive to indoors and target areas.	[29, 45, 74]	Bluetooth and motion sensor were used together. A self-correction system was developed for evaluation of the different results obtained from the two methods [29].
Infrared	It is a data collection technology that can also be used indoors without lighting.	[33, 76, 78, 79]	In a dark environment, a system has been proposed that determine the characteristics of the targeted object with artificial neural networks, and then increases the performance of the tracking with a correlation filter [44].
Ultrasound	Ultrasound technology is a technology that determines the location of an object by extracting sound waves that the human ear can not hear.	[34, 46, 79, 80]	Ultrasound technology was used to develop a system that detects their proximity to specific locations in the museum by sending sound waves to mobile phones of users [34].

an integrated circuit tag to a host computer with a wireless exchange [52]. The first application areas of RFID can be listed as the retail sector, supply chain management, logistics management, manufacturing and military applications [53]. Apart from these, there are also applications such as hospitals [54, 55], museums [56] and shops [22]. There are two types of RFID tags, active and passive. Active tags have a small power source that is used to send a signal, which is in a range of up to 100 milliseconds, while passive tags do not have a power source and are usually activated by a signal at a shorter distance than one meter. A typical RFID system consists of three parts: sensor, reader and control application. Sensors on the person or object usually contain a unique code that is used to identify a tracking object. When the object reader is within an appropriate range, the sensors send the data to the reader using radio signals. The reader converts the radio signal to the digital information that is then routed to a computer application that uses it [57].

WiFi is one of the most popular and widely used wireless technologies. Many indoor facilities such as hospital [58], university [59, 60] and indoor [61] are used for wireless Internet connection. It is a great advantage to be found on a wide range of devices such as mobile phones, notebooks, PCs and tablets and to use WiFi technology widely used by users. Each WiFi card allows measuring some signal parameters available for the indoor locations [62]. WiFi technology uses signals to detect the presence of a device. The system determines the location of the device based on signals from other access points [63]. One advantage of the indoor positioning is that power consumption is usually very low due to the short distance. Also, the accuracy levels can be quite high, but it depends on the system used [57]. On the other hand, the use of WiFi signals can cause problems with absorption and reflection. The required WiFi parameters are not always available in modern telephone operating systems [46, 64].

Bluetooth is also a very popular wireless technology. It is the most used technology for devices such as laptops, tablets, smartphones and even smart clocks. Bluetooth tracking method uses wireless Bluetooth signal [44, 65]. Bluetooth scanners are installed in places of interest. These scanners scan Bluetooth signals emitted by mobile phones and record every detection. Unlike RFID, there is no need to pre-register the tracking object with Bluetooth and add any device or tag [44, 66]. The advantage of Bluetooth is that it removes the worry that people will be followed, making the collected data natural and error-free. Compared to WiFi, Bluetooth uses shorter range, smaller bandwidth but much more power [62, 67]. There are also some situations where validation is satisfactory in indoor locations since Bluetooth does not affect user behavior [23, 68].

ZigBee is the best technology for low power, short range (max 100 m) and relatively low data flow rate (up to 100kbps) [69]. The ZigBee network layer consists of star, tree and mesh networks. In a star network, there is a coordinator that allows devices connected to it. The ZigBee coordinator is responsible for initiating, maintaining and monitoring the network. In tree and network networks, devices can communicate with each other in a multi-hop fashion. The network has a ZigBee coordinator and multiple ZigBee routers. A device can join a network as a router or an end device by associating it with a coordinator or a parent router. In the tree network, coordinators and routers can periodically announce the signal lamps to start the superframes [25].

Computer vision technology, which collects data via image recording, uses advanced image processing algorithms with the complex structure to describe the object to be tracked [23]. Although the camera collects very rich data about explored environment [70, 71], detection-based viewers have poor performance because the detection of objects is not reliable [72, 73]. Also, if more detail about the behavior of the customers is needed, the operators are sometimes asked to directly observe the behavior of the customers in the indoor location [16].

Motion sensors are usually equipped with three types of sensors: a gyroscope, an accelerometer, and a magnetometer. In general, the gyroscope and accelerometer combine data to determine the orientation and location of a monitored body part [74], and magnetometers are sensors that are used to calibrate the direction of the sensor [29].

The infrared (IR) wavelength is longer than the visible light length but shorter than the trillion hertz (TeraHertz) radiation [75, 76]. For this reason, infrared light is invisible to the human eye in most conditions. With this feature, infrared technology helps to gather data without intervention by the user for indoor positioning [77, 78].

Ultrasound technology is a technology that uses ultrasonic frequencies near the human perception threshold (around 20 kHz) to describe the existence of an object [35]].

Ultrasound technology is better at making fast and accurate detection of objects than motion detection sensors [46, 79]. However, user posture and clothing may affect to measure distances or may mislead or distort ultra-sonic bumps. This affects the performance of ultra-sound sensors negatively [79].

2.3 Evaluation Criteria

The criteria for selection of the most appropriate IBAT are determined by an extended literature review and expert opinion. The decision model for the selection problem is given in Figure 2.1. The model is composed of four levels. At the first level the main goal, which is to select the best IBAT takes place. At the second level, four main criteria namely, economic factors (C_1), data accuracy (C_2), data quality (C_3) and technical factors (C_4) take place. Economic factors represent the monetary value of the investment, and it is composed of two sub-criteria: cost (C_{11}) and time (C_{12}). Data accuracy criteria focus on the accuracy of the data captured by the IBAT alternative, and it involves two sub-criteria accuracies (C_{21}) and bias (C_{22}). Data quality criteria focus on the capacity of the collected data, and it involves quantity (C_{31}) and introspection (C_{32}) sub-criteria. The final main criteria are the technical factors which focus on simplicity (C_{41}) of the IBAT alternative and ease of analysis (C_{42}) of the data collected by the alternative.

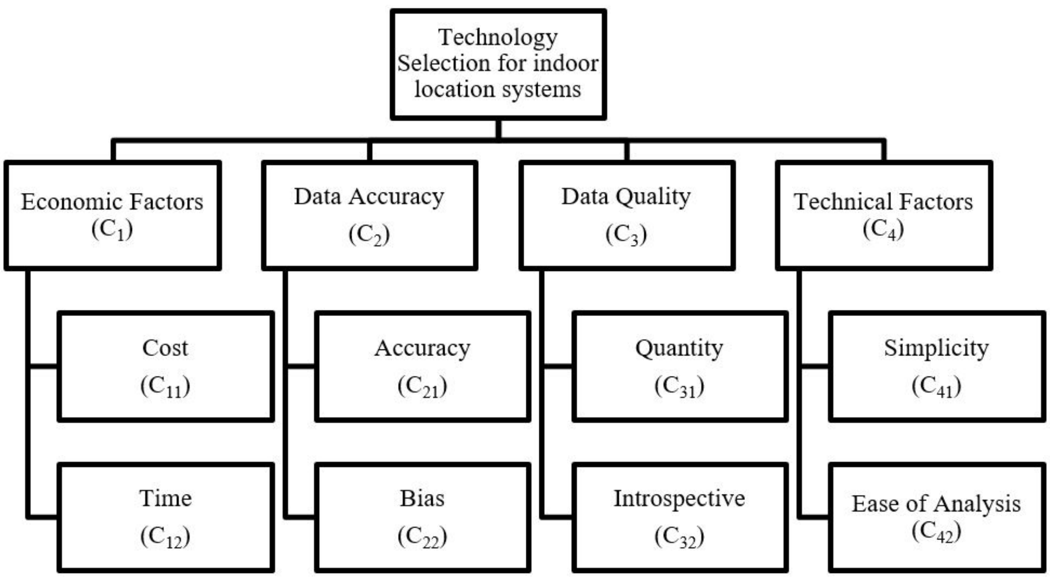

Figure 2.1 : Hierarchy of Criteria.

.

The details about the sub-criteria are as follows:

Cost (C_{11}): It refers to the total cost required to collect the data. This criterion is considered as the cost of the human resource used in the data collection, the material source such as the machine or device to be purchased and all other expenses.

Time (C_{12}): It refers to the time required to collect and analyze data. It is expected to be evaluated as high (H) if the collection and analysis period is low.

Accuracy (C_{21}): It refers to the degree of accuracy of aggregated data. Therefore, it is also a criterion used to see how close obtained the results are to the actual method used.

Bias (C_{22}): It refers to the degree of error as a result of the data analysis. The results of the analysis are biased in cases where the data contain noise values, extreme values, and inappropriate analytical values.

Quantity (C_{31}): According to Frisby et al. [80] "when we increase the scale of the data that we work with, we can do new things that were not possible when we just worked with smaller amounts." So the more data is collected, the closer the result will be. The quantity criterion refers to the number of data collected in an equal duration for each method.

Introspection (C_{32}): It expresses the ability to obtain information that reflects the internal world, such as customers' goals, satisfaction, experience, expectation. In some cases, customer considerations become important, in addition to the meaning derived from the data. Therefore, the introspective feature of the data is a criterion to consider.

Simplicity (C_{41}): It refers to the effort to collect data. The simplicity criterion indicates both the ability to collect data and the ease of preparation, which is necessary before the data is collected For example the method can require the establishment of a system or if the method is one of the traditional methods, training of interviewers, preparation of questions may be necessary.

Ease of Analysis (C_{42}): It refers to the easy assessment of collected data. In some cases, the data must be preprocessed or processed with various algorithms before analysis. Because these situations make the analysis difficult, ease of analysis is a criterion that should be considered in the selection of the data collection method.

2.4 Methodology

2.4.1 Preliminaries

Let U be the universe of discourse, $U = [0, u]$. A trapezoidal fuzzy number (TrFN) can be defined as $\widetilde{A} = (a_l, a_{m1}, a_{m2}, a_u)$, where $0 \leq a_l \leq a_{m1} \leq a_{m2} \leq a_u$ as shown in Figure 2.2.

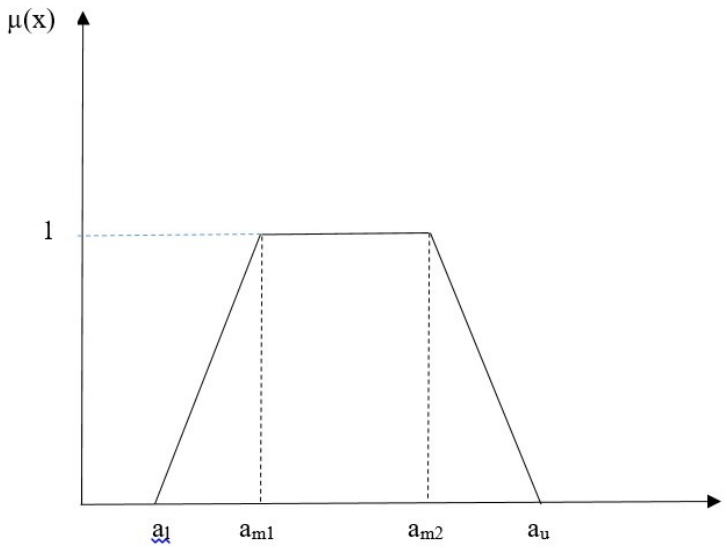

Figure 2.2 : Membership function of trapezoidal fuzzy number.

The membership function of $\widetilde{A} = (a, b, c, d)$ is presented as in Equation (2.1).

$$\mu_{\widetilde{A}}(x) = \begin{cases} \frac{x-a_l}{a_{m1}-a_l}, \; for \; a_l \leq x \leq a_{m1} \\ 1, \; for \; a_{m1} \leq x \leq a_{m2} \\ \frac{a_u-x}{a_u-a_{m2}}, \; for \; a_{m2} \leq x \leq a_u \\ 0, \; for \; otherwise \end{cases} \tag{2.1}$$

Let $\widetilde{A} = (a_l, a_{m1}, a_{m2}, a_u)$ and $\widetilde{B} = (b_l, b_{m1}, b_{m2}, b_u)$ be two TrFN, some of the arithmetic operations are as follows [81]:

Addition:

$$\widetilde{A} \oplus \widetilde{B} = (a_l + b_l, a_{m1} + b_{m1}, a_{m2} + b_{m2}, a_u + b_u) \tag{2.2}$$

Multiplication:

$$\widetilde{A} \otimes \widetilde{B} = (a_l \times b_l, a_{m1} \times b_{m1}, a_{m2} \times a_{m2}, a_u \times b_u) \tag{2.3}$$

Multiplication with a crisp value (k):

$$k \times \widetilde{A} = (k \times a_l, k \times a_{m1}, k \times a_{m2}, k \times a_u) \tag{2.4}$$

Division:

$$\widetilde{A} \oslash \widetilde{B} = (a_l \setminus b_u, a_{m1} \setminus b_{m2}, a_{m2} \setminus a_{m1}, a_u \setminus b_l) \tag{2.5}$$

*n*th power:

$$\widetilde{A}^n = (a_l^n, a_{m1}^n, a_{m2}^n, a_u^n) \tag{2.6}$$

Aggregation of fuzzy numbers is very important for multi-expert decision making. Yager [82] proposed ordered weighted averaging (OWA) given in Equation 2.7. An OWA operator of dimension n is a mapping $F : Rn \to R$, that has an associated n vector

$$OWA(a_1, a_2, \cdots, a_n) = \sum_{j=1}^{n} w_j \times b_j \tag{2.7}$$

where b_j is the jth largest element or the set and $w_j in[0,1]$ and

$$\sum_{i=1}^{n} w_i = w_1 + w_2 + \cdots + w_n = 1 \tag{2.8}$$

Filev and Yager [83] define the weight vector of OWA operators $W^1 = (w_1^1, w_2^1, \cdots, w_n^1)$ and $W^2 = (w_1^2, w_2^2, \cdots, w_n^2)$ as given in Equation (2.9) and Equation (2.10).

$$w_1^1 = \alpha_2, w_2^1 = \alpha_2(1-\alpha_2), \cdots, w_{n-1}^1 = \alpha_2(1-\alpha_2)^{n-2}, w_n^1 = (1-\alpha_2)^{n-1}. \tag{2.9}$$

$$w_1^2 = \alpha_1^{n-1}, w_2^2 = (1-\alpha_1)\alpha_1^{n-2}, \cdots, w_n^2 = 1-\alpha_1 \tag{2.10}$$

Recently, new extensions of ordinary fuzzy sets have been proposed to involve decision makers' hesitancy into the decision model. One of the most recent fuzzy set extensions is Hesitant Fuzzy Sets (HFSs). HFSs have been introduced by Torra [37] for cases where defining the membership degree of an element is not easy since several possible membership values may exist.

Let X be a fixed set, a hesitant fuzzy set (HFS) on X is regarding a function that when applied to X returns a subset of [0, 1] [37]. The mathematical expression for HFS is as follows:

$$E = < x, h_E(x) > | x \in X \tag{2.11}$$

where $h_E(x)$ is a set of some values in $[0,1]$, denoting the possible membership degrees of the element $x \in X$ to the set E.

Based on HFS, Rodriguez et al. [38] introduced the concept of a hesitant fuzzy linguistic term set (HFLTS) to provide a linguistic and computational basis to increase the richness of linguistic elicitation based on context-free grammars by using comparative terms.

An HFLTS is defined as an ordered finite subset of consecutive linguistic terms of $S = s_0, \cdots, s_g$. For instance, let S be defined as S = { s_0: nothing, s_1 : very bad, s_2 : bad, s_3 : medium, $s = 4$: good, $s = 5$: very good, $s = 6$: perfect } and v be a linguistic variable, $H_s(v) = bad, medium, good$.

2.4.2 Steps of Hesitant Fuzzy Analytic Hierarchy Process

The analytic hierarchy process, developed by Saaty [36], is an MCDM structured technique based on pairwise comparisons, for solving complex decision-making problems. In traditional AHP expert evaluations are expressed by using crisp numbers. To handle the vagueness and imprecision in experts' evaluations, fuzzy AHP methods have been proposed [84–86]. Fuzzy AHP methods employ triangular or trapezoidal fuzzy numbers to represent linguistic terms. Later AHP method has also been extended using extensions of fuzzy sets [87–89]. Recently Hesitant Fuzzy AHP is proposed which is based on HFLTS definition [39, 40].

In this study, we modify existing Hesitant Fuzzy AHP method to employ trapezoidal fuzzy numbers instead of fuzzy triangular numbers. The steps of the proposed hesitant fuzzy AHP are as follows:

Step 1. The decision makers decide on the alternatives and form the decision model.

Step 2. Decision makers evaluate criteria, sub-criteria and alternatives using HFLTS and the context-free grammar "between" and "is". Decision makers use the linguistic scale given in Table 2.2 to fill the pairwise comparison matrices.

Step 3. Form the aggregated pairwise comparison matrix $\widetilde{P}$.

Table 2.2 : Linguistic scale for hesitant fuzzy AHP.

	Linguistic Term	Abbreviation	TrFN
s10	Absolutely High Importance	(AHI)	(7,8,9,9)
s9	Very High Importance	(VHI)	(5,6,8,9)
s8	Essentially High Importance	(ESHI)	(3,4,6,7)
s7	Weakly High Importance	(WHI)	(1,2,4,5)
s6	Equally High Importance	(EHI)	(1,1,1,3)
s5	Exactly Equal	(EE)	(1,1,1,1)
s4	Equally Low Importance	(ELI)	(0.33,1,1,1)
s3	Weakly Low Importance	(WLI)	(0.2,0.25,0.50,1)
s2	Essentially Low Importance	(ESLI)	(0.14,0.2,0.25,0.33)
s1	Very Low Importance	(VLI)	(0.11,0.13,0.17,0.2)
s0	Absolutely Low Importance	(ALI)	(0.11,0.11,0.13,0.14)

$$\widetilde{P} = \begin{vmatrix} \widetilde{a}_{11} & \widetilde{a}_{12} & \cdots & \widetilde{a}_{1n} \\ \widetilde{a}_{21} & \widetilde{a}_{22} & \cdots & \widetilde{a}_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \widetilde{a}_{n1} & \widetilde{a}_{n2} & \cdots & \widetilde{a}_{nn} \end{vmatrix} \tag{2.12}$$

where $\widetilde{a}_{ij}$ represents the aggregated evaluation on comparison of the ith element to jth element. In the aggregation process, the decision makers make their evaluations separately, if the evaluation on a specific comparison is not close, the decision makers review their evaluations. If the evaluations are the same or close to each other then they are represented using "is" and "between" keywords.

Step 4. Transform $\widetilde{P}$ matrix into numerical pairwise comparison matrix ($\widetilde{C}$) by using OWA operations given in Equation (2.9) and Equation (2.10) [38].

$$\widetilde{C} = \begin{vmatrix} (1,1,1,1) & \widetilde{c}_{12} & \cdots & \widetilde{c}_{1n} \\ \widetilde{c}_{21} & (1,1,1,1) & \cdots & \widetilde{c}_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \widetilde{c}_{n1} & \widetilde{c}_{n2} & \cdots & (1,1,1,1) \end{vmatrix} \tag{2.13}$$

where $\widetilde{c}_{ij} = (c_{ij_l}, c_{ij_{m1}}, c_{ij_{m2}}, c_{ij_u})$. The reciprocal of $\widetilde{c}_{ij}$ is obtained as follows:

$$\widetilde{c}_{ij} = (\frac{1}{c_{ij_u}}, \frac{1}{c_{ij_{m2}}}, \frac{1}{c_{ij_{m1}}}, \frac{1}{c_{ij_l}}) \tag{2.14}$$

If all decision makers use the same linguistic for an evaluation, then the associated trapezoidal fuzzy number given in Table 2.2 is used. If the resulting evaluation contains "between" keyword, such as "between s_t and s_x" where the scale contains eleven terms and $s_0 \leq s_t \leq s_x \leq s_10$, then the parameters of $\widetilde{c}_{ij}$ is determined using Equations (2.15) – (2.18).

$$c_l = min\{c_l^t, c_{m1}^t, c_{m2}^t, c_l^{i+1}, \dots c_{m1}^x, c_{m2}^x, c_u^x\} = c_l^t \tag{2.15}$$

$$c_r = max\{c_l^t, c_{m1}^t, c_{m2}^t, c_l^{i+1}, \dots c_{m1}^x, c_{m2}^x, c_u^x\} = c_u^x \tag{2.16}$$

$$c_{m1} = \begin{cases} c_{m1}^t, \ if\ t+1 = x \\ OWA_{w^2}\left(c_{m1}^t, \cdots, c_{m1}^{\frac{t+x}{2}}\right), \ if\ t+x\ is\ even \\ OWA_{w^2}\left(c_{m1}^t, \cdots, c_{m1}^{\frac{t+x+1}{2}}\right), \ if\ t+x\ is \quad odd \end{cases} \tag{2.17}$$

$$c_{m2} = \begin{cases} c_m^t + 1, \ if\ t+1 = x \\ OWA_{w^1}\left(c_m^x, c_m^{x-1}, \cdots, c_m^{\frac{t+x}{2}}\right), \ if\ t+x\ is\ even \\ OWA_{w^1}\left(c_m^x, c_m^{x-1}, \cdots, c_m^{\frac{t+x+1}{2}}\right), \ if\ t+x\, is\ odd \end{cases} \tag{2.18}$$

The weights used in the OWA operator are calculated based on α_1 and α_2 values. In our case, these parameters are calculated as $\alpha_1 = \frac{g-(x-t)}{g-1}$, $\alpha_2 = \frac{(x-t)-1}{g-1}$, where g is the number of terms in the evaluation scale.

Step 5. Calculate the geometric mean for each row ($\widetilde{r}_i$) in $\widetilde{C}$ using Equation 2.19:

$$\widetilde{r}_i = (\widetilde{c}_{i1} \otimes \widetilde{c}_{i2} \otimes \cdots \otimes \widetilde{c}_{in})^{1/n} \tag{2.19}$$

Step 6. Calculate the fuzzy weight ($\widetilde{w}_i$) for each row using Equation (2.20):

$$\widetilde{w}_i = \widetilde{r}_i \otimes (\widetilde{r}_1 \oplus \widetilde{r}_2 \cdots \oplus \widetilde{r}_n)^{-1} \tag{2.20}$$

Step 7. Repeat the steps 2-6 for each level of the decision-making model to find the local weights of main-attributes, sub-attributes, and the alternatives.

Step 8. Calculate the global weight of each sub-attribute, by multiplying its local weight with the weight of the associated attribute. Then, calculate the fuzzy performance score of each alternative by Equation (2.21).

$$\widetilde{S}_i = \sum_{j=1}^{n} \widetilde{w}_j \otimes \widetilde{s}_j, \ \forall i. \tag{2.21}$$

Step 9. Compare the fuzzy performance score of the alternatives by using Equations (2.22) – (2.25) defined by Kumar et al. [90].

$$If\ R(\widetilde{c}_1) > R(\widetilde{c}_2)\ then\ \widetilde{c}_1 > \widetilde{c}_2\ and\ If\ R(\widetilde{c}_1) = R(\widetilde{c}_2)\ then\ \widetilde{c}_1 = \widetilde{c}_2 \tag{2.22}$$

$$IfRM(\widetilde{c}_1) > RM(\widetilde{c}_2)\ then\ \widetilde{c}_1 > \widetilde{c}_2 \tag{2.23}$$

$$R(\widetilde{c}) = \frac{\widetilde{c}_1 + \widetilde{c}_{m1} + \widetilde{c}_{m2} + \widetilde{c}_u}{4} \tag{2.24}$$

$$RM(\widetilde{c}) = \frac{\widetilde{c}_{m1} + \widetilde{c}_{m2}}{2} \tag{2.25}$$

2.5 Application

We apply the steps of the proposed hesitant fuzzy AHP method in the following.

Step 1: consider a technology selection problem with four alternatives (RFID, WiFi, Bluetooth and camera) and four main criteria, each of them has two sub-criteria. The hierarchy of the problem is given in Figure 2.1. The considered main criteria are economic factors (C_1), data accuracy (C_2), data quality (C_3) and technical factors (C_4). The sub-criteria are cost ($C_1 1$), time ($C_1 2$), accuracy ($C_2 1$), bias ($C_2 2$), quantity ($C_3 1$), introspective ($C_3 2$), simplicity ($C_4 1$), ease of analysis ($C_4 2$), respectively. The abbreviations of the linguistic variables are represented in Table 2.2.

Step 2 and 3: Tables 2.3 and 2.4 present pairwise comparison matrices for criteria and alternatives using linguistic terms (w.r.t.: with respect to).

Table 2.3 : Pairwise comparison matrix for criteria.

	C_1	C_2	C_3	C_4
C_1	1	VLI-WLI	EsLI-EqLI	WHI-EsHI
C_2		1	EqHI-WHI	VHI-AHI
C_3			1	EsHI-VHI
C_4				1

Table 2.4 : Pairwise comparison matrix for alternatives.

	w.r.t. C_1			w.r.t. C_2	
	A1	A2		A1	A2
A1	1	WHI-EsHI	A1	1	EqHI-WHI
A2		1	A2		1
	w.r.t. C_3			w.r.t. C_4	
	A1	A2		A1	A2
A1	1	WLI-EqLI	A1	1	VHI-EsHI
A2		1	A2		1

Step 4: Linguistic evaluations are transformed to numerical evaluations using OWA operator. Tables 2.5 and 2.6 indicate the transformed evaluations of Tables 2.3 and 2.4.

Table 2.5 : Transformed numerical values of pairwise comparison matrix for criteria.

	C_1	C_2	C_3	C_4
C_1	(1,1,1,1)	(0.11,0.13,0.28,1)	(0.14,0.18,0.55,1)	(1,2,6,7)
C_2	(1,3.64,7.74,9)	(1,1,1,1)	(1,1,4,5)	(5,6,9,9)
C_3	(1,1.82,5.71,7)	(0.2,0.25,1,1)	(1,1,1,1)	(3,4,8,9)
C_4	(0.14,0.17,0.5,1)	(0.11,0.11,0.17,0.2)	(0.11,0.13,0.25,0.33)	(1,1,1,1)

Table 2.6 : Transformed numerical values of pairwise comparison matrix for alternatives.

	w.r.t. C_1		w.r.t. C_2	
	A1	A2	A1	A2
A1	(1,1,1,1)	(1,2,6,7)	A1 (1,1,1,1)	(1,1,4,5)
A2	(0.14,0.17,0.5,1)	(1,1,1,1)	A2 (0.33,1,1,1)	(1,1,1,1)
	w.r.t. C_3		w.r.t. C_4	
	A1	A2	A1	A2
A1	(1,1,1,1)	(0.2,0.25,1,1)	A1 (1,1,1,1)	(3,4,8,9)
A2	(0.33,1,1,1)	(1,1,1,1)	A2 (0.33,1,1,1)	(1,1,1,1)

For an example calculation, consider evaluation of VLI–WLI comparison for C_1 and C_2 in Table 2.3. VLI and WLI correspond {0.11, 0.13, 0.17, 0.20} and {0.20, 0.25, 0.50, 1}, respectively. For left and right TrFN are calculated using Equations (2.6) and (2.7), respectively.

$c_l = min\{0.11, 0.13, 0.17, 0.20, 0.20, 0.25, 0.50, 1.00\} = 0.11$

$c_r = max\{0.11, 0.13, 0.17, 0.20, 0.20, 0.25, 0.50, 1.00\} = 1.00$

Middle values m_1 and m_2 are calculated according to Equation (2.17) and Equation (2.18), respectively. Since, i equals to 1 for VLI and j equals to 3 for WLI, the result of $i+j$ is even. Therefore, $c_{m1} = OWA_{w^2}(a_m^i, \cdots, a_m^{i+j/2})$ is used. To calculate c_{m1}, we need to use Equation (2.5). Equation (2.10) is used to calculate the second kind of weights, w_1^2. Since $\alpha_1 = (11-(3-1))/(11-1) = 0.9$ and $\alpha_2 = ((3-1)-1)/(11-1) = 0.1$, $c_{m1} = 0.90.13+0.10.2 = 0.13$. Similar calculations are made for c_{m2} , $c_{m2} = 0.10.50+0.90.25 = 0.28$

Step 5. The matrices are defuzzified and examined for consistency.

Step 6. The geometric mean of each row is calculated using Equation (2.19). The geometric mean of the first row ($\widetilde{r}_1$), which belongs to C_1 in Table 2.5.

$\widetilde{r}_1 = [(1,1,1,1)\otimes(0.11,0.13,0.28,1.00)\otimes(0.14,0.18,0.55,1.00)\otimes(1,2,6,7)]^{1/4}$
$= (0.35,0.46,0,98,1.63)$

Step 7. The fuzzy weight ($\widetilde{w}_i$) is calculated by using Equation (2.20).

$\widetilde{w}_i = (0.35,0.46,0,98,1.63) \otimes ((0.35,0.46,0,98,1.63) \oplus (1.5,2.16,4.09,4.49) \oplus (0.88,1.16,2.6,2.82) \oplus (0.2,0.22,0.38,0.51))^{-1}$
$= (0.04,0.06,0.24,0.55)$

Step 8. The steps 2-6 are repeated for each level of the decision-making model.

The local weights of the criteria and sub-criteria are shown in Table 2.7 and Table 2.8.

Step 9. By multiplying criteria local weights with sub-criteria local weights, the global weights of the sub-criteria are obtained. Later,the local weights of the alternatives are multiplied with the associated sub-criteria global weights of the alternatives (Table 2.9).

Table 2.7 : Local weights of the criteria.

C_1 (0.04,0.06,0.24,0.55)	C_{11} (0.27, 0.45,1.34,1.92)	C_{31} (0.22,0.25,0.67,0.98)
C_2 (0.16,0.27,1.02,1.53)	C_{12} (0.1,0.13,0.39,0.73)	C_{32} (0.29,0.5,0.67,0.98)
C_3 (0.09,0.1,0.65,0.96)	C_{21} (0.31,0.33,1.0,1.42)	C_{41} (0.43,0.52,0.94,1.3)
C_4 (0.02,0.03,0.09,0.17)	C_{22} (0.18,0.33,0.5,0.63)	C_{42} (0.14,0.26,0.33,0.43)

Step 10. The fuzzy performance score of each alternative is compared, and the results are shown in Table 2.10.

According to R values shown in Table 2.10 (R_4 >R_1 >R_2 R_3), the best alternative is chosen as A4 (Bluetooth). Under this expert evaluations Bluetooth is followed by A1 (Wifi), A2 (Camera) and A3 (RFID), respectively.

Table 2.8 : Local weights of the sub-criteria.

w.r.t. C_{11}	w.r.t. C_{12}	w.r.t. C_{21}
A1 (0.08,0.11,0.48,0.95)	A1 (0.07,0.08,0.11,0.25)	A1 (0.17,0.27,0.68,1.02)
A2 (0.04,0.06,0.27,0.6)	A2 (0.04,0.08,0.11,0.14)	A2 (0.04,0.06,0.18,0.34)
A3 (0.02,0.03,0.09,0.21)	A3 (0.06,0.08,0.11,0.19)	A3 (0.13,0.22,0.63,0.9)
A4 (0.14,0.32,1.08,1.8)	A4 (0.42,0.54,0.93,1.11)	A4 (0.05,0.06,0.15,0.34)
w.r.t. C_{22}	w.r.t. C_{31}	w.r.t. C_{32}
A1 (0.06,0.12,0.46,0.89)	A1 (0.09,0.12,0.92,1.15)	A1 (0.08,0.09,0.15,0.19)
A2 (0.04,0.05,0.13,0.36)	A2 (0.08,0.1,1.36,1.88)	A2 (0.08,0.09,0.15,0.19)
A3 (0.02,0.04,0.12,0.21)	A3 (0.04,0.06,0.43,0.55)	A3 (0.08,0.09,0.15,0.19)
A4 (0.21,0.37,1.04,1.54)	A4 (0.02,0.03,0.39,0.59)	A4 (0.3,0.41,1.03,1.3)
w.r.t. C_{41}	w.r.t. C_{42}	
A1 (0.05,0.07,0.22,0.38)	A1 (0.14,0.2,0.52,1.08)	
A2 (0.09,0.13,0.5,0.82)	A2 (0.11,0.2,0.39,0.72)	
A3 (0.21,0.33,1.02,1.42)	A3 (0.07,0.15,0.3,0.37)	
A4 (0.02,0.03,0.07,0.11)	A4 (0.06,0.11,0.3,0.37)	

Table 2.9 : Global weights of the alternatives.

wrt C_{11}	wrt C_{12}	wrt C_{21}
A1 (0,0,0.16,1.01)	A1 (0,0,0.01,0.1)	A1 (0.01,0.02,0.69,2.21)
A2 (0,0,0.09,0.64)	A2 (0,0,0.01,0.06)	A2 (0,0.01,0.18,0.74)
A3 (0,0,0.03,0.23)	A3 (0,0,0.01,0.08)	A3 (0.01,0.02,0.64,1.95)
A4 (0,0.01,0.35,1.91)	A4 (0,0,0.09,0.45)	A4 (0,0.01,0.16,0.74)
wrt C_{22}	wrt C_{31}	wrt C_{32}
A1 (0,0.01,0.23,0.86)	A1 (0,0,0.4,1.08)	A1 (0,0.01,0.07,0.18)
A2 (0,0,0.06,0.35)	A2 (0,0,0.59,1.76)	A2 (0,0.01,0.07,0.18)
A3 (0,0,0.06,0.21)	A3 (0,0,0.19,0.52)	A3 (0,0.01,0.07,0.18)
A4 (0.01,0.03,0.53,1.5)	A4 (0,0,0.17,0.55)	A4 (0.01,0.03,0.45,1.21)
wrt C_{41}	wrt C_{42}	
A1 (0,0,0.02,0.09)	A1 (0,0,0.02,0.08)	
A2 (0,0,0.04,0.18)	A2 (0,0,0.01,0.05)	
A3 (0,0,0.09,0.32)	A3 (0,0,0.01,0.03)	
A4 (0,0,0.01,0.03)	A4 (0,0,0.01,0.03)	

Table 2.10 : Comparison for each alternative.

	Fuzzy Weights	R	RM
A1	(0.02, 0.05, 1.59, 5.61)	1.81896	0.82297
A2	(0.01, 0.03, 1.06, 3.95)	1.26155	0.5419
A3	(0.01, 0.04, 1.09, 3.5)	1.16029	0.56626
A4	(0.02, 0.08, 1.76, 6.4)	2.06686	0.92112

2.6 Conclusion

The emerging technologies provide various alternatives capturing indoor behaviors which lead to technology selection problem for companies. In this paper, the IBAT selection problem is modeled as an MCDM containing four criteria and eight sub-criteria. The contribution of the paper is two folds. First, the IBAT selection problem has been defined and solved using Hesitant Analytic Hierarchy Process for the first time. Also, the original Hesitant AHP methodology is extended by using trapezoidal fuzzy numbers to incorporate expert evaluations in a better way. For further research, other fuzzy AHP methods like Laarhoven and Pedrycz's [84] may be extended by hesitant fuzzy sets and the obtained results can be compared with ours. Other multi-attribute decision-making techniques such as; ELECTRE, VIKOR, and MACBETH can also be extended by using HFS and HFLTS.

3. ANALYSIS OF FREQUENT VISITOR PATTERNS IN A SHOPPING MALL[2]

Current technological developments provide a huge amount of data to collect, store and process. One of the study topics which use collected data is human behavior analysis. Companies try to generate customer-oriented solutions by analyzing their interests and desires. For online companies, it is easier because of the simplicity of data collection. Each click or cursor movement is recorded easily. On the other hand, physical retailers have a limitation that is the collection of customer data. Infosys Company was reported that 78% of consumers would be customers again if a retailer offers customer-oriented solutions [17]. In the light of this study, understanding customer is a critical factor for not only online companies but also physical retailers to survive in a competitive environment.

RetailNext [20] declares that according to 84% of customers, retailers should integrate the online and offline channels. Therefore, physical stores need to provide customer-based service that they received from the Internet by establishing relationship digital channels [21]. Tracking customers in the stores using different technologies such as Bluetooth facilitates integrating online and offline channels [16, 21–23]. The increase in mobile device use of customers brings an opportunity for retailers to develop customer experience in-store shopping [91]. Bluetooth technology may be used to track customers' digital footprints. These digital footprints can be used to detect the customers' movement in-store and common routes. The results of the analysis are also used for in-store marketing purposes.

In this study, we use data collected via Beacon devices. The iBeacon devices (hereafter Beacons) broadcast low energy radio signals, interact with the smart devices in their coverage zone, and exchange information with them. The coverage zone of a Beacon device is 50 meters, and their interaction distance can be fine-tuned

[2]This chapter is based on the paper Dogan, O., Gurcan, O.F., Oztaysi, B. and Gokdere, U. (2019). Analysis of Frequent Visitor Patterns in a Shopping Mall. In *Industrial Engineering in the Big Data Era*, (Ed.s) Calisir, F., Cevikcan, E. and Camgoz Akdag, H. Springer Berlin Heidelberg, pp. 217–227.

within this range. This interaction, which is extremely power-efficient, is based on Bluetooth's innovative network technology Bluetooth Low Energy (BLE), or the trademarked name, Bluetooth® Smart. When a mobile device encounters a Beacon device in a venue, the mobile device automatically refers to a software program (SDK) embedded into a mobile application running on the mobile device. Consequently, it consults to the beacon's server. According to the command from the server, the mobile device performs any action, such as displaying a message or serving as a micro-location sensor. The beacon device, the embedded SDK in the application, and the server, together form the sensor management platform of Blesh that realizes the Beacon communication. The data used in this study is collected by Blesh, a Turkish start-up company which offers enterprises the whole platform along with the assisting service to manage, focusing on refining the opportunities that come along with this platform, and how this new communication method can produce business solutions. As Blesh beacons, mobile application partners and the enterprise network grow, the beacon management platform turns gradually into an advertising platform for marketing agencies to get involved in. Exceeding 100.000 beacons in over 50 cities in Turkey, amounting to roughly 15.000 points of interest, Blesh beacons are placed in more than 200 shopping malls, prominent airports, leading banks, top retail shops, subway stations, cafes and restaurants and among many other venues.

3.1 Literature Review

In the literature and industry, various ways are used to collect the data for customer behavior analysis. As a traditional way, the questionnaire is the most commonly preferred method to collect data among others such as observation and interview. Das and Varshneya [92] and Kesari and Atulkar [93] used a structured questionnaire to understand consumers' emotions and satisfaction, respectively, in a shopping mall. Some researches use together questionnaire and review methods to enhance significance. For example, Kirchberg and Tröndle [41] used the interview method to research the similarities and differences between their experiences. The collected data are grouped into three categories: visitors' expectations refer to the motivations before the visit, experiences indicate social interactions during the visit and contributions point out what customers learned after the visit. Yiu and Ng [94] also used together with customer surveys and interview methods to measure buyers-to-shoppers ratio. Merad et al. [16] used an observation method in their study. Operators have directly observed customers' behaviors in the store.

As an alternative to data collection to analyze customer behavior, some technologies gather much more data in a non-invasive way. Traditional data collection methods, questionnaire, interview, observation or surveys, have some limitations. For instance, real-time of customers' activities, generally, cannot be saved. This results in deviations from actual results. For that reason, Willeims et al. [95] used WiFi technology to compile an inventory of retail technologies, which are classified based on shopping value types and shopping cycle stage. Carrera et al. [96] used WiFi technology for real-time tracking. Fukuzaki et al. [97] try to estimate the actual number of people in the area by comparing the number of acknowledged devices in the shopping mall by using WiFi. Oosterlinck et al. [23] demonstrated the applicability of Bluetooth tracking in a shopping mall. They used Bluetooth scanners in their study. They obtained high data quality and low cost of data collection. Yewatkar et al. [98] proposed a smart shopping cart system that will keep track of purchased products and online transaction for billing using RFID and ZigBee. Although Zigbee and Bluetooth technologies collect data in a similar way, the difference between them is the data exchange scale. Zigbee gathers data on a small scale.

The camera is another data collection technology. Camera technologies use sophisticated image processing algorithms that are complex to describe the object to be tracked [23]. Wu et al. [21] used a camera to record customers in the store to understand their behavior. They showed the trajectories of the customer in a heat map. Arroyo et al. [99] used the in-store camera to detect real-time detection of suspicious customer behaviors in a shopping mall. They benefited from advanced processing algorithms in their study. Table 3.1 shows some studies used data collection technologies.

3.2 Methodology

Association Rule Mining (ARM) is one of the data mining methods used to discover interesting dependencies between datasets [106]. Many business enterprises collect a huge amount of data from various sources. Table 3.2 shows an example of data related

Table 3.1 : Some data collection technologies and studies.

Technology	Study	Technology	Study
Bluetooth	[23] [44] [67] [68] [80] [100]	Zigbee	[25] [69] [101]
RFID	[28] [30] [54] [98] [102]	Camera	[21] [26] [31] [73] [99]
Wi-Fi	[43] [95] [96] [97] [103] [104]	Infrared	[33] [78]
		Ultrasound	[34] [105]

to customer routes in a shopping mall. Each location in the shopping mall has a beacon ID that collects data from customers' mobile phone via Bluetooth.

Table 3.2 : An example of customer routes.

Customer ID	Routes
4633698	{121600, 121633, 121565}
26066031	{121372, 121600, 121307, 121507}
29235005	{121289, 123937, 121388, 121372}
37711500	{121372, 121289, 121600, 121388}
35272768	{121372, 121289}

Agrawal et al. [107] first introduced ARM as an effective way to find relationships between hidden large datasets that satisfy certain support and confidence values. Interesting relations or correlations are helpful in many businesses for making useful decisions [108]. Therefore, many industrial applications try to find dependencies between variables.

Support and confidence values are two critical parameters in ARM. If the support value is very low then generated rule may occur by chance. Also, very low support threshold may cause to generate uninteresting rules [109]. Frequent item set generation is a result of the support threshold. On the other hand, confidence check reliability of the inference made by rule. Minimum confidence value affects the generated rules. Equation (3.1) and Equation (3.2) show the calculation of the support and confidence ratios, respectively.

$$Support, s(X \rightarrow Y) = \frac{n(X \bigcup Y)}{N} \tag{3.1}$$

$$Confidence, c(X \rightarrow Y) = \frac{n(X \bigcup Y)}{n(X)} \tag{3.2}$$

where N is a total number of transactions, X and Y show two items in the dataset.

An association rule is created by item sets that satisfy the minimum support ($minsup$) and minimum confidence ($minconf$) values. It does not imply causality. Causality needs cause and consequent of the rule. Table 3.3 demonstrates the interestingness of a rule in terms of support and confidence. An interesting rule is a rule that has lower support value and higher confidence value.

Table 3.3 : Interestingness of a generated rule.

		Support	
		Low	High
Confidence	Low	Uninteresting	Uninteresting
	High	Interesting	Uninteresting

ARM represents the uncovered relationships as sets of frequent items or association rules. Frequent item sets generation is the first step of ARM. It aims to find items or item sets that satisfy the minimum support constraint. Since the generation of frequent item sets has computational requirements, it is more expensive than those rule generation. Therefore, some effective methods were developed. Aggrawal and Srikant [107] presented the Apriori algorithm to facilitate frequent item set generation. Apriori algorithm helps to reduce the number of candidate item sets to generate frequent item sets. Han et al. [110] developed the FP-Growth (Frequent Pattern Growth) algorithm to discover frequent item sets without any item set generation. As the second step, rule generation is used to find item sets that satisfy the minimum support value and a predetermined confidence value.

A rule extracted from the large dataset is shown $X \rightarrow Y$. For example, the rule $\{121372\} \rightarrow \{121289\}$ suggests that there is an important and strong relationship between location 121372 and location 121289. This means customers who visit the location 121372 strongly visits the location 121289. Finding these types of rules has an opportunity to increase cross-selling by understanding customer needs.

Under normal circumstances, k-items-dataset can generate $2^k - 1$ candidate frequent itemset, excluding the null set. Also, the total number of rules that can be extracted form dataset is . More than 80% of the rules are not useful when $minsup$ is 20%, and $minconf$ is 50%, thus making most of the computations become wasted [109]. Since many applications have a very large k value, the number of itemsets increase exponentially. Frequent itemsets are determined to reduce wasted computations. Therefore, two algorithms to avoid that problem in frequent itemset generation step are developed. That is: (i) reduce the number of candidate itemsets (Apriori algorithm) and (ii) reduce the number of comparisons (FP-Growth algorithm).

Apriori algorithm helps to reduce the number of candidate itemsets to generate frequent itemsets. Apriori principle is based on that if an itemset is frequent, then all of its subsets must also be frequent [109]. This principle known as support-based pruning has a key property, monotonicity property, considering support value. According to the monotonicity property, an itemset never exceed the support value of its subset. In the Apriori algorithm, an anti-monotone itemset that not satisfy monotonicity property is used to reduce candidate itemsets.

FP-Growth algorithm does not need any candidate itemset like Apriori algorithm. It uses a compact data structure, named FP-tree to encode the data and generates frequent itemsets directly from this structure [109].

3.3 Application Data

3.3.1 Data Collection and Preparation

The data used in this study is collected from Beacons located in one of the major shopping malls in Istanbul. Beacons broadcast low energy radio signals, interact with the smart devices in their coverage zone, and automatically exchange information with them if the user allows. The data used in this study is collected by Blesh, a Turkish start-up company which has already located the beacon devices to the shopping mall. The system can recognize each smart device by its unique id. Besides, each beacon device has a specific id, which is associated with a single location, in our case retail store. The data is composed of monthly transactions of 12070 unique users in the shopping mall with 293 retail stores. Since the coverage zone of a beacon device is 50 meters, it is possible that a smart device can get signals from two or more beacons at the same time. In this case, the approximate distance between the smart device and each beacon is calculated. During the preparation phase, the signals from the closest location are preserved the other signals are deleted.

3.3.2 Results

Association rule mining approach is used for the analysis of the visitor data. As the first step, the frequent item sets are prepared. We used a minimum support value of 0.05, and no frequent sets are observed having more than seven elements. We observed only one set which has seven elements and 100 sets which have six elements. The results show that the most commonly used routes have at most seven stores at a confidence level of 0.01.

When we analyze the stores that take place in frequent item sets, we can observe only 19 stores while there are 293 in total. Also, we calculate the number of total observations for each store in the frequent item set. The results reveal that the top seven "most frequent" stores cover more than 80 percent of the total observations.

Table 3.4 shows the stores that take place in frequent paths. Due to data security issues, only the IDs of the retail stores can be published. We also calculate the popularity of the stores by summing up the total signals and rank the stores based on their popularity. It is interesting that Store 121599, which is the most popular store, takes the fourteenth place in the share. In the other direction, Stores 121532 and 123881 are most frequent stores. However, they are eighth and seventh in popularity. Besides, among the top ten most popular stores only six of them takes place in the frequent store's list. This shows that frequent store definition, which is based on association rule mining is different from popular store definition, which is calculated by overall total signals is different.

Table 3.4 : Stores in the frequent paths and their popularity ranking.

Store	Share	Popularity Rank	Store	Share	Popularity Rank
121532	15.70%	8	121373	1.70%	18
123881	14.70%	7	121466	1.50%	19
121485	12.10%	2	121324	1.50%	55
121506	10.50%	17	121599	0.80%	1
121378	10.30%	21	121294	0.80%	20
121570	9.40%	15	121563	0.50%	3
121472	8.20%	12	121549	0.50%	4
121545	6.00%	13	123807	0.30%	114
121481	2.80%	38	121495	0.20%	30
121382	2.60%	88			

Analysis of the association rules also reveals interesting results. When minsup is 0.05 and minconf is 0.1, 13735 rules are generated. Some of the most interesting rules, which has lower support and higher confidence ratio, are shown in Table 3.5.

Table 3.5 : Generated interesting rules.

Premises	Conclusions	Support	Confidence
{123881, 121373, 121324}	121485	0.05	0.92
{121563, 121324, 121368}	121485	0.05	0.89
{123881, 121570, 121439}	121532	0.05	0.86
{121532, 121506, 121324}	123881	0.05	0.85

According to Table 3.5, although visiting the locations 123881, 121373, 121324 and 121485 is a rare route, it has high confidence. This means that the number of four locations visited together is low (only 5% of all routes), but when the premises locations are visited together, it is very likely (92%) that the next visit will be the location 121485 (conclusion) for customers who visit the locations 123881, 121373 and 121324. As an example, the rule is shown by $\{123881, 121373, 121324\} \rightarrow \{121485\}$.

3.4 Discussion and Conclusion

In this paper, we analyze the visiting data of a shopping mall collected via Beacon devices. Basically, heat maps are prepared based on this type of data, and they show the popularity of locations. In other terms, they graphically demonstrate how many times visitors exist at that location. In this study, we also obtain the popularity of the stores and the most frequent routes. The results reveal that the locations on frequent routes are not always popular. In other terms, frequent routes are different from popular locations. Thus, further studies can be conducted to improve analytic results using route information.

On the other hand, the representative power of the data is a limitation of the study. Although smart device penetration is high, people may choose not to activate their Bluetooth receiver. Thus, the collected data may not represent all visitors of the shopping mall. Data collected using other similar technologies, such as video recording and Wi-Fi may also be used to reach enhanced results.

4. ANALYZING OF GENDER BEHAVIORS FROM PATHS USING PROCESS MINING: A SHOPPING MALL APPLICATION[3]

Customer loyalty has critical importance for business continuity. The underlying problem is to understand customers. There are some researches on the relationship between customer royalty and individualized proposals. For example, Infosys Company says companies that offer customer–oriented solutions have a big advantage concerning customer royalty that 78% of consumer will be again customer [17]. As another example, Microsoft asserts that 73% of customers prefer companies that offer individualized purchasing solutions [18]. As a brand separator, customer experience will be more important than product and price by 2020 [19]. The research indicates the importance of understanding today's customers having a digital purchasing experience.

Although individualized and customer–oriented solutions are more manageable for online companies, it is more difficult for the physical store because of data collection about customers. On the other hand, according to the research by RetailNext [20], 84% of customers think that retailers should integrate the online and offline purchasing channels. Behavioral analytics can integrate online and offline channels for physical stores using different human tracking technologies [16, 21–23]. RFID (Radio Frequency Identification), WiFi (wireless fidelity), Bluetooth, ZigBee are alternative technologies for customer tracking data. The technology is an essential factor affecting customer relationship in physical stores. Since the customer behavior analysis (CBA) needs to data gathered by unobtrusively, the data should be captured using one of the indoor location system (ILS) sensors. In our previous study, we studied on which data collection technologies is more appropriate for indoor customer tracking using fuzzy multi-criteria decision-making technique [14]. As a result, we installed iBeacon devices, Bluetooth-based data collection technology, into the shopping mall. Gathering data from customer's smartphone connecting with Bluetooth provides the location and timestamp data belonging to the customer. Information can be generated on where

[3]This chapter is based on the paper Dogan, O. Bayo-Monton, J.L., Fernandez-Llatas, C. and Oztaysi, B. (2019) Analyzing of Gender Behaviors from Paths Using Process Mining: A Shopping Mall Application. *Sensors 19*(3), 557–577.

the customers are in the shopping mall, the time they spend in each visited store and the next visit, etc. The created information also can be used for sales and marketing purposes [13, 14].

Studying to understand indoor customer behavior concerning pathways is an interesting area [111]. In this study, we divide customer data as man and woman customers to investigate whether there is a behavioral difference between man and woman customers. We selected a shopping mall in Istanbul as an indoor area because retail sector employees are under high pressure due to inadequate professional training and human resources, a large number of operations, poor management and poor customer-oriented service [112]. For that reason, better and more efficient use of resources is necessary for the improvement of the retail system and for strengthening customer relations. Performing simple statistics do not give general information about customers. One problem is that a global view of customers' pathway to make clear to understand by humans [42]. Besides, it is a requirement not only to discover a pathway that followed by customers but also to create an understandable output. Both discovery of pathways for a general view and generating understandable results are one of the goals of process mining.

Process mining (PM) is a technique that aims to extract understandable information via process models generated automatically by different discovery algorithms from event logs using process management and business analytics [11, 12]. A primary event log contains a case ID, activity and timestamp data. The more data attributes such as resource, cost, etc., the detailed information about the process. A process model shows the flow of the data visually to make understandable the process. There can be different types of notations such as Petri Nets, BPMN (Business Process Modeling Notation) and EPC (Event–driven Process Chains). The process model is the output of the process discovery. Process mining has some discovery algorithms to generate automatically process models such as alpha algorithm [113], fuzzy mining [114], heuristic mining [115], genetic mining [116], social network mining [117], inductive mining [13] and PALIA Suite (Parallel Activity Log Inference Algorithm) [55].

Most of the pathway analyses use graphical representation to show the pathways. Without any detailed study, the created pathways are intuitive [13]. However, process mining explains the underlying information of the pathway. Therefore, process mining is selected to discover pathways of indoor customers to understand the behavior of customers visiting the shopping mall. In this study, behavioral analysis refers to understand customer needs and propose the best solutions to the needs. The study does not aim to investigate the behavioral science related to retail.

This chapter is structured as follows. Firstly, studies related to customer behavior analysis and process mining are examined to show the importance of this study. Secondly, the selected algorithm and the proposed system are introduced in detail. Thirdly, the experimental case study is shown. Then, the discussion and conclusion section is presented. Finally, the limitations of the study are discussed.

4.1 Related Works

There are many studies in the literature on behavior analysis to better understand people. It has many indoor implementations such as museum [44], hospital [118] [119], shopping [99,120], and airport [121]. The researchers collected data in different ways. For example, Yoshimura et al. [44] took advantage of Bluetooth, on the other hand, Arroyo et al. [99] gathered data by recording with the camera.

Yoshimura et al. [44] used Bluetooth to determine the duration of the visiting of visitors in Louvre Museum by taking into consideration the way of walking and the layout of the museum. As a result, museum visitors were categorized according to their duration of stay. Oosterlinck et al. [23] showed possibilities of usage of Bluetooth tracking in a shopping mall. They asserted Bluetooth has a high data quality and low cost of data collection. Although camera technologies need complex image processing algorithms [23] and have poor performance because detecting people occurrences is not reliable [72], it is an excellent alternative to record every view about customers. Merad et al. [16] performed purchasing behavior analysis according to the intensities of computer vision technology and customers in the store. Different algorithms were used to track the people in the video recordings. Wu et al. [21] created pathways of the customer in a retail store using a heat map.

Process mining has increasing and attractive importance in the literature. Efficiently collecting data through a wide variety of sensors provides for extracting valuable information about processes. Since healthcare has very dynamic and sophisticated process [119], many researchers concentrate on different healthcare processes. Rovani et al. [122] showed the differences between actual processes and guidelines described concerning the best-practices using conformance checking in the urology department. Mans et al. [123] applied process mining to dentistry. They first discovered the process model with heuristic mining algorithm and then enhanced the discovered process. Fernandez-Llatas and colleagues [42] presented a process mining based method using indoor location systems. With this method, data was gathered with RFID technology, without forcing nurses to measure time, and other information about the processes and enter into the systems. The data was used as input for the process mining and the flows of the different processes were obtained. In another study, Fernández-Llatas et al. [58] used process mining to analyze the movements of people with a 25-week data belonging to 9 people and collected by wireless technology.

Process mining has been applied to not only healthcare but also municipality [11,124], informatics [125–127], education [128–130], finance [131, 132], and manufacturing [113,133] .

We have realized from the literature review that data collected by different technologies are analyzed for various purposes. Table 4.1 shows some of them.

Table 4.1 : Purpose of the collected data.

Purposes	Studies
To determine some parameters (average number of places visited by people, the average time spent by them, most visited / least visited places) with descriptive statistics	[44]
To discover routes followed by customers	[21,44]
To estimate the next visiting place	[31]
To determine where the customer is located at any time	[23,72]

In our study, we used the collected data with Bluetooth to determine some descriptive statistics and discover pathways of the customers in the shopping mall.

Table 4.2 gives information on some studies. PM column shows the studies collecting data with one of the technologies and creating pathways but not focusing on customer behaviors. The study considers papers that collect data with one of the wireless or computer vision technologies such as RFID, WiFi, Bluetooth, camera in the literature review. Some paper used data obtained from IT systems such as ERP [134] or social media such as Instagram [135]. Although these studies are not directly related to CBA, we include them in the literature review because they have valuable results.

Table 4.2 : Summary of literature review.

Study	CBA	Technology	PM	Implementation Area
[136]	✓	Bluetooth		Museum
[44]	✓	Bluetooth		Museum
[68]	✓	Bluetooth		Exhibition
[23]	✓	Bluetooth		Store/Shopping Mall
[80]	✓	Bluetooth		Hospital
[137]	✓	Bluetooth		Museum
[100]	✓	Bluetooth		Museum
[21]	✓	Camera		Store/Shopping Mall
[99]	✓	Camera		Store/Shopping Mall
[120]	✓	Camera		Store/Shopping Mall
[15]	✓	Camera		Store/Shopping Mall
[16]	✓	Camera		Store/Shopping Mall
[54]		RFID		Hospital
[138]	✓	RFID		Store/Shopping Mall
[139]	✓	RFID		Store/Shopping Mall
[22]	✓	RFID		Store/Shopping Mall
[13]	✓	WiFi	✓	Store/Shopping Mall
[58]	✓	RFID	✓	Hospital
[135]	✓	Other	✓	Exhibition
[134]	✓	Other		Manufacturing

There is only one paper focusing on customer behavior analysis using process mining and collecting data with a wireless sensor in an indoor location. Hwang and Jang [13] published a study that shows the possibilities of analyzing customer behavior using WiFi-based technology and the process mining technique. They used WiFi technology to gather customer data in a retail store in South Korea and applied inductive mining, one of the process mining discovery tools, to create customers trajectories. Although inductive mining removes infrequent activities and paths [140], it has some disadvantages. When the event log is incomplete or has exceptional data, the inductive miner is relatively sensitive [140]. In such cases, inductive miner algorithms cannot detect a strong relation in the event log. The second significant disadvantage of inductive mining is that it cannot deal with duplicate activities in the event log [141]. In some cases to discover shoppers' paths, activities with the same label can be observed in more than one location because beacons are generally located near to each other in an indoor location.

4.1.1 Original Aspect of The Study

Although, both PALIA Suite and inductive miner algorithms produce process trees, PALIA Suite has extra steps to tackle duplicate activities in the event log. It can combine nodes according to the name or importance. Importance between two nodes refers that the node names are the same, but there is a difference between them concerning spent time in the node, PALIA Suite proposes two option: combine whatever it is or not combine if it has a significant difference. On the other hand, inductive mining algorithm aims to remove infrequent activities in the log to extract meaningful information. However, infrequent information can carry critical information about daily movements of customers. For example, a visit to the cinema can be only once in a month and according to inductive mining view, this data should be removed from dataset. Nevertheless, this visit may be critical for specifying a behavioral change of the customer. For example, not visiting the cinema can be a withdrawal symptom of timelessness or economic break. PALIA Suite can consider infrequent behavior with statistical information about the activities frequencies and state changes [58].

One of the main problems of the application of Process Mining Techniques in environments with high variability is the appearance of the undesirable Spaghetti Effect Problem [142]. The Spaghetti Effect is one of the most challenging problems in process mining studies [11]. It is related to the difficulty of readability of sophisticated processes. Many transitions presented in the process model look like a spaghetti. The high variability produced by the human behavior makes results in a complex understanding of the flows inferred due to the appearance of all the possible movements that the users have. Also, the objective of this work is not only to support professionals in a better understanding of the human behaviours at malls, but also to identify the different behaviours of different kind of users. In recent works [143], clustering techniques are applied to discover groups of behaviours These techniques are based on the application of classic clustering algorithms such as k-means or quality threshold cluster, in combination with workflow distance measures [58] that groups the most common behaviours. This combination allows, on the one hand, to group common traces that have a similar behavior, and, on the other hand, limit the Spaghetti Effect because of the elimination of the variability in each one of the groups. PALIA Suite has some clustering techniques implemented, allowing the discovery of different groups in an easy way. This tool has been used in real scenarios [143].

The other difference of our study from the study of Hwang and Jang [13] is the application area. Collecting data in only one retail store is more accessible than collecting in a shopping mall. At the same time, a shopping mall case requires more time than a store case for preprocessing steps and analyzing data. They collected data in a store, on the other hand, we collected data in one of the biggest shopping malls in Turkey, which has six floors and 293 stores.

On the other hand, selection of the data collection technology is also essential. When we compare WiFi and Bluetooth, Bluetooth technology outperforms WiFi technology concerning simplicity and accuracy criteria while they have the similar levels for the time, cost and bias [14]. At the same time, WiFi signals may have a problem with absorption [144]. Collecting comprehensive indoor data with WiFi is a big problem [64]. Bluetooth is a cost effective and unannounced indoor tracking technology that provides unbiased and uninfluenced observations [68]. Also, Bluetooth indoor tracking provides fast and accurate solutions [23]. According to Oosterlinck and friends [23], Bluetooth is a better alternative than WiFi because Bluetooth is more specifically designed for short ranges in indoor locations. Moreover, Bluetooth gives higher accuracy for an indoor location implementation. In summary, Bluetooth is a better data collection technology than WiFi in terms of indoor localization studies.

4.2 Methodology

For taking advantage of the aggregated data, new technologies are needed to inform the whole process, making it easier for people to understand process characteristics. Process mining (PM) is a new way of using event logs to extract models that people can read [145]. Process mining uses event logs to discover, monitor and improve processes. Event logs are the sum of the events considered as input for the process mining. In the Industrial 4.0 Era, the digital world and the physical world are now intertwined. Today's information systems hold event logs in extreme quantities. On the other hand, today's technologies enable to collect much more data from smart devices. These opportunities make process mining a preferred technique to analyze collected data.

Process mining has three kind of study topics: Discovery, Conformance checking and Enhancement [11]. The process discovery is the most important output of a process mining study [124]. Process discovery techniques use event logs to generate a process model without using any prior knowledge. The discovered model is represented by different notations such as Petri Net and BPMN, which describes the behavior in the logs. There are several discovery algorithms such as heuristic mining [146], genetic mining [147], alpha [12] and PALIA Suite [148]. In this study, we used PALIA Suite [58].

PALIA Suite is a discovery algorithm creating graph-based process models [42]. ILS data analysis is one of the implementation areas of graph-based process models that aims to present sophisticated information in a readable and understandable way. [42]. PALIA is based on Grammar Inference Theory and in this case, it can be seen as a kind of Bayesian Network with the enhancement. Besides, PALIA keeps the traceability of all the real events that are related to each one of the nodes and transitions. Therefore, after discovery, researchers can recover the set of events related to each one of the nodes and transitions, and even, create new statistics. It consists of five main steps [55]. Figure 4.1 shows the steps of the PALIA Suite.

Figure 4.1 : PALIA Suite steps.

- *The Parallel Acceptor Tree Algorithm* step produces process trees to start and end time of events and their parallelism.

- *Onward Merge* step fuses all branches of the tree. Algorithm checks for each node that following two branches are equivalent. If so, they are fused. If all nodes and transitions use the same token for the same process in the two branches, it is said two branches are equivalent.

- *The Parallel Merge* step merges nodes that are sequential and represent the same event.

- *Delete Repeated Transitions* step deletes repeated transitions.

- *Delete Unused Nodes* step deletes unused nodes.

Finally, the algorithm produces a process model described as timed parallel automatons (TPA) that models the system. TPA represents a workflow as expressive as Petri nets [149].

For this study, we gathered data from customers visiting a shopping mall in Istanbul. The collected data contained infrequent paths typical of customer activity. For this reason, we needed a process mining algorithm that could cope with an infrequent and incomplete event log. PALIA Suite can consider infrequent behavior with statistical information about the activities frequencies and state changes [58]. In the case of the Clustering algorithm we selected a Quality Threshold Cluster algorithm [150] that allows group by a similarity between the traces using a Workflow distance [58] based on Error Correcting paradigm [151] for limiting the Spaghetti Effect and Maximizing the similarity among the traces.

4.3 iBeacon ILS

With the spread of the internet of things, Beacon is a generic name given to small Bluetooth radio transmitters that offer personalized experiences. Beacon technology provides location information using Bluetooth low energy (BLE) or the trademarked name, Bluetooth® Smart. Each iBeacon device has a coverage zone at 50–meter distance, and the interaction distance among devices can be fine-tuned in the zone. For that reason, one smart device can receive signals from more than one beacon devices. In this case, proximity distances between the device and beacon are calculated and classified as near, medium and far.

For the study, we collected data via iBeacon devices located to the shopping mall by a beacon network company, Blesh Company (Blesh.com) in Turkey. Approximately 300 iBeacon devices are located in the shopping mall for 293 stores by Blesh Company. Figure 4.2 depicts the main architecture of Beacon technology and communication around this architecture. As a beacon network company, Blesh provides a software development kit (SDK) to third part application developers. Various third part mobile applications use this SDK to communicate with the beacon network [152]. By this way, the applications can provide an enhanced communication and value added services

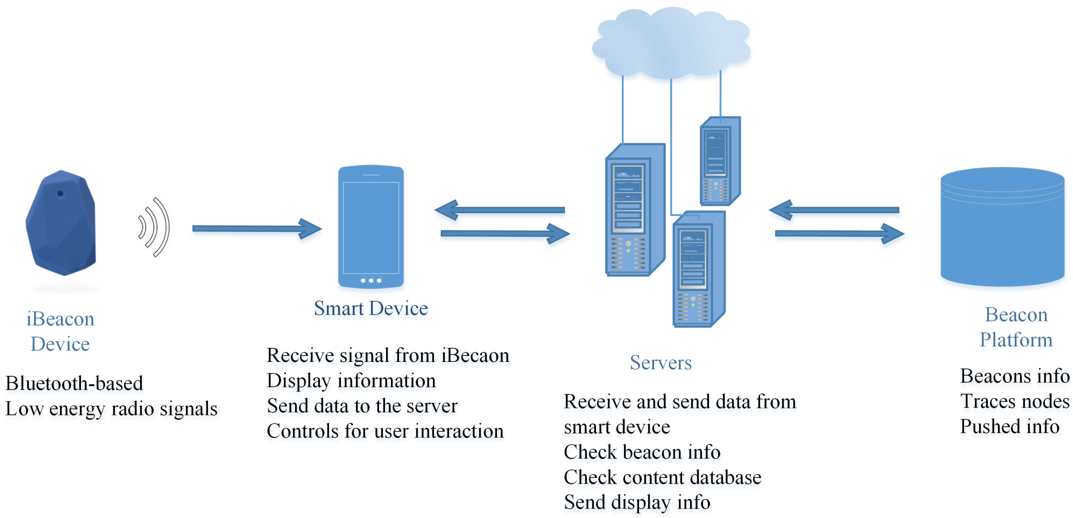

Figure 4.2 : The architecture of Beacon technology.

to their customers through their own interface, after getting permission from the customers.

iBeacon devices transmit low energy radio signals and interact with the smart devices in their coverage zone. If the user allows, they automatically save the customer data including a unique ID and time data. Each iBeacon device has a unique ID that represents a store. The system recognizes each smart device by its unique ID. Therefore, it allows us to create a pathway of the customer in the shopping mall. The Beacon platform includes all beacons and location data. The server receives data from the smart device and contacts with the Beacon platform. Then it sends display information to the smart device.

iBeacon ILS can produce CSV files that have in each line a positioning value for each event. Table 4.3 shows a sample part of the localization events provided by Beacons. In each line, the following data are available:

- *ID*: For each captured data, a unique number is defined to use different analysis. ID column is not used in our study.
- *Dongle Columns*: Because, in some cases, iBeacon devices located near each other, one customer can be seen in different three locations. According to the proximity of the customer to the stores, iBeacon devices gather three different position data. Dongle_1 shows the nearest store that customer data captured, on the other hand,

Table 4.3 : An example of an event log created from iBeacon ILS.

ID	Dongle_1	Dongle_2	Dongle_3	Timestamp	SubscriberID
1028333326	121527			11.12.2017 00:14:42	17399446
1028334382	121498			11.12.2017 00:16:48	39081930
1028334406	121498	121404		11.12.2017 00:16:50	39081930
1028334421	121498			11.12.2017 00:16:53	39081930
1028492822	121436	121510	121446	11.12.2017 07:23:47	29078632
1028492925	121510	121372		11.12.2017 07:23:59	29078632
1028492939	121436	121510	121446	11.12.2017 07:24:01	29078632
1028495185	121446	121436	121510	11.12.2017 07:28:23	29078632

Dongle_3 represents the farthest store location. In the study, we ignore Dongle_2 and Dongle_3 localization data.

- *Timestamp*: This is a date and time value that represents the moment at which data are captured by the iBeacon device for the related store. Timestamp format is dd.mm.yyyy hh:mm:ss.

- *SubscriberID*: This number shows the customer identification number associated with the mobile.

Four basic perspectives are used in process mining [11] using the data attributes. *Control-flow perspective* considers the order of activities. The purpose of mining in this perspective is to find a good flow that represents all possible ways. *The organizational perspective* focuses on resources such as people, systems, roles or departments in event records. Their relation to each other is examined. The aim is to show the social relations between resources or to classify resources regarding organizational unit and role. *Event perspective* (Case perspective) focuses on the properties of events. An event can be represented by the process or by the initial activity of the event. For example, in a material replacement order, it may be important to know the supplier or know the number of products ordered. *The time perspective* concerns the time or frequency of events. When the start and end times of the events are known, it is possible to determine the bottlenecks in the process, to measure the level of service, to view the use of resources, and to predict the remaining process.

4.4 Experimental Results

The data used in this research is collected in one of the major shopping malls in Istanbul. The shopping mall has six floors with a total of 293 shops. At the beginning of the study, 779877 raw data were collected which belong to 12000 unique customers. We grouped the stores concerning their products or services. For example, Store A, Store B and Store C is grouped into Clothing shop group, if they sell mainly clothes. After some other data preparations, we used 642 customer data that we can determine genders. Table 4.4 shows customer data details regarding three months. In the collected data, there are 1293 cases that show the number of customer paths in each visit. In these cases, 2749 store groups are visited. During these three months, at least one customer visited the shopping mall 52 times in a month. Because of the filter that we applied named as "Shop Split Filter by Gap", the number of visits is higher than the total number of days in the month.

Table 4.4 : Summary of corpus after combining to shop groups.

	Customer Data
Total Customer (Man/Woman)	642 (165/477)
Total Number of Cases	1293
Maximum Number of Visit Sessions	52
Localization Events	2749

We hypothesize that the use of process mining algorithms provides understanding indoor customer behaviors from their pathways in a human–understandable view and a behavioral difference based on gender. A detailed review of how customer behavior change for behavioral science is not in the scope of this paper. One of the aims of this study is to show the potential of the combination of iBeacon ILS with process mining techniques.

For performing a primary process mining application, an event identification, a timestamp, and an activity attributes are required from the event log. In our study to enhance the analysis, some additional data attributes are added to corpus such as day number, day, day group, floor and time interval.

4.4.1 Data Preparation

To apply process mining with PALIA Suite, firstly, we created an event log from collected ILS data. CustomerID, a unique identifier (ID), was created to track every customer. Location shows the stores in the shopping mall and is used to create customers' pathways. ILS dataset includes a timestamp data. Instead of a timestamp, PALIA Suite requires a start time and an end time for each localization data. Therefore, we prepared our data by creating a start and end time from the timestamp. One customer can visit the shopping mall in different sessions. For example, customer X visits the shopping mall in the morning between 10.00 and 10.50. In the same day, the customer again visits the shopping mall from 19.30 to 21.00. These two visits are considered different sessions because the study aims to determine customer behavior. The customer may have different purposes during his/her visits. Moreover, since store group data is more meaningful and easy to interpret regarding the aim of the study, we also combine stores into store groups.

We applied some filters to extract more information. For example, the working time restricted between 10:00:00 and 22:00:00 because of the opening time of the shopping mall. Shop Duration Filter adjusts that the visit duration must be more than 1 minute. Otherwise, it is considered that the customer data is captured while walking instead of visiting. Therefore, a time gap occurs. Let the path (Store1 – 11:10:00), (Store2 – 11:14:00), (Store3 – 11:15:00), (Store1 – 11:22:00) shows an event log belonging to the same customer. We can easily say that the customer visited Store1 at 11:10:00 and left at 11:13:59. Then, (s)he went to the Store2. When we delete the Store2 event due to 1-minute filter, we encounter a time gap problem. The customer was in Store1 until 11:13:59 and then seen in Store3 at 11:15:00. Where is the customer between 11:13:59 and 11:15:00? This is a kind of missing data. After Shop Duration Filter, we fuse consecutive customer data called as Fuse Consecutive Events Filter to avoid this disappearance. Fuse Consecutive Events Filter adjusts the ending time of Store1 visiting as 11:14:59, one second before the starting of the consequent event. Shop Split Filter by Gap is applied if the time gap in between two subsequent observations in a path has a duration of 90 minutes, it is considered as a new visit. If an event is the last event in the trace, its end time is determined by adding 1 minute to its start time. Only the signals from the nearest location detected by iBeacon devices are assumed as the location where the customer is. The other signals are ignored and can be considered as missing data. In that case, we use Dongle_1 Location Data.

4.4.2 Process Mining

Although basic statistics do not give results on the general view of the customer pathways, they are useful to see some data-centric results.

Table 4.5 : Basic statistics of the preprocessed data.

	December 2017	January 2018	February 2018
Total Customer (Man/Woman)	290 (89/201)	181 (47/134)	171 (29/142)
Total Number of Cases	450	444	399
Maximum Number of Visit Sessions	45	52	43
Localization Events	957	1088	704

Table 4.5 represents the size of data after preprocessing steps. For example, during the last month of 2017, we collected the 290 customer data, 89 man data and 201 woman data. In our study, we assume that if one customer has a 90-minute interval between two consecutive events, it is considered as a different visit. Therefore, during December, we have 450 different cases that refer to different visits by 290 customers. There is at least one customer that have 45 visit sessions.

Figure 3 shows that there is a decrease in the last week of the month regarding the number of customers. Although, the number of customers in December is more than the number of customers in January, customers in January visited more store groups than the other two months. At the same time, Figure 4.3 shows some of the customers visited more than one time on the same day, resulting in inequality between the number of customers and number of cases.

Although collected data do not contain gender information, it is clear that customer going to Toilette Woman is a woman and customer going to Toilette Man is a man. In the figures showing customer paths, nodes and arrows are colored from green to red. The red nodes refer the store group that has higher case duration. On the other hand,

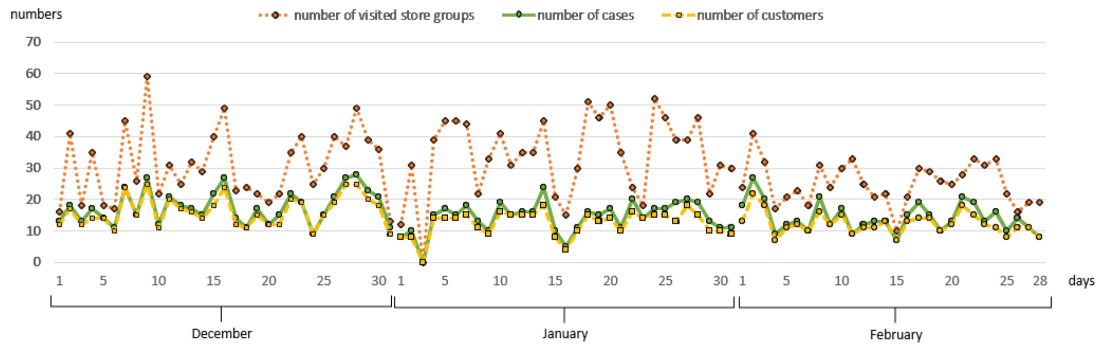

Figure 4.3 : Number of customers and cases.

the red arrows present the mostly executed transaction between two nodes concerning the number of execution.

Since we cluster the customer paths with PALIA Suite to find meaningful and readable results, the path graphs do not represent all data in the event log. The graphs represent all man and woman paths with rates 0.824, 0.803 in December 0.811, 0.835 in January and 0.861, 0.824 in February, respectively man and woman. In the 4-faced figures, (a) and (c) refer to mostly followed man customer paths and (b) and (d) refer to mostly followed woman customer paths. For example, Figure 4.5 includes two paths mostly followed by man and woman customers in December. A sample data is given in Figure 4.4 to validate for Figures 4.5, 4.6 and 4.7.

Man Paths

Customer_ID	Store_Group	Start_Time	End_Time
23359612	Clothing	15.12.2017 14:48	15.12.2017 15:58
23359612	Supermarket	15.12.2017 15:59	15.12.2017 16:12
23359612	Clothing	15.12.2017 16:13	15.12.2017 16:28
24271462	Accessory	31.12.2017 14:15	31.12.2017 14:29
24271462	Catering	31.12.2017 14:30	31.12.2017 15:50
34692929	Catering	7.01.2018 14:53	7.01.2018 15:54
34692929	Clothing	7.01.2018 15:55	7.01.2018 16:44
34692929	Catering	7.01.2018 16:45	7.01.2018 16:51
35045082	Clothing	11.01.2018 13:14	11.01.2018 14:10
35045082	Accessory	11.01.2018 14:11	11.01.2018 14:19
35045082	Clothing	11.01.2018 14:20	11.01.2018 14:33
35045082	Catering	11.01.2018 16:55	11.01.2018 17:12
35045082	Clothing	12.01.2018 17:13	12.01.2018 18:11
35045082	Supermarket	12.01.2018 18:12	12.01.2018 18:24
35045082	Clothing	12.01.2018 18:25	12.01.2018 18:46
37781110	Clothing	21.02.2018 20:09	21.02.2018 20:16
37781110	Catering	21.02.2018 20:17	21.02.2018 20:57
37600854	Catering	5.02.2018 10:22	5.02.2018 11:24
37600854	Home	5.02.2018 11:25	5.02.2018 11:45

Woman Paths

Customer_ID	Store_Group	Start_Time	End_Time
16507273	Clothing	7.12.2017 17:19	7.12.2017 18:53
16507273	Catering	7.12.2017 18:54	7.12.2017 19:22
16507273	Clothing	7.12.2017 19:23	7.12.2017 19:47
16507273	Home	7.12.2017 19:48	7.12.2017 19:58
16507273	Electronics	7.12.2017 19:59	7.12.2017 20:09
9048839	Entertainment	2.12.2017 11:15	2.12.2017 11:23
9048839	Clothing	2.12.2017 11:24	2.12.2017 11:31
9048839	Catering	2.12.2017 11:32	2.12.2017 11:54
39329211	Clothing	2.01.2018 12:11	2.01.2018 12:36
39329211	Accessory	2.01.2018 12:37	2.01.2018 12:39
39329211	Clothing	2.01.2018 12:40	2.01.2018 13:01
39329211	Accessory	2.01.2018 13:02	2.01.2018 13:40
23115102	Entertainment	2.01.2018 13:41	2.01.2018 15:46
23115102	Catering	2.01.2018 15:47	2.01.2018 16:21
38822141	Accessory	21.02.2018 11:09	21.02.2018 11:54
38822141	Clothing	21.02.2018 11:55	21.02.2018 12:02
38822141	Accessory	21.02.2018 12:03	21.02.2018 12:23
38822141	Clothing	21.02.2018 12:24	21.02.2018 12:29
38822141	Catering	21.02.2018 12:30	21.02.2018 12:52

Figure 4.4 : Sample validation data for discovered customer paths.

In December, Catering and Clothing are the most popular store groups in terms of the number of visits and duration (Figure 4.5). The numbers of customers who visit Catering then Clothing and vice versa they are 16 and 14 in man customer visits and 35 and 47 in woman customer visits, respectively. Man customers visit fewer store groups than woman customers. Although Accessory is a common store group, woman customers spend less time. Moreover, Entertainment and Personal Care have significance concerning the behavioral difference. Man customers do not visit any Entertainment and Personal Care stores in mostly followed paths.

When we compare December and January, woman customers in January prefer Entertainment instead of Clothing (Figure 4.6). The Entertainment duration was increased in January because some of the highly anticipated movies came out. Man customers have still visited Catering and Clothing. Besides, Electronics and Mother&Baby have more duration in man customer visits. Home is another difference from December for woman customers. The number of visiting Home is nearly similar, but the duration in Home is four times larger than duration in December. It is interesting that, although woman customers in Figure 4.6(b) mostly visit Clothing, they spend more time in Entertainment. In many cases, customers visiting Entertainment visit a Catering store to eat or drink.

Figure 4.7 represents February customer paths. Man customers mostly visit Electronics and Catering store groups. On the other hand, Clothing and Accessory store groups have the highest duration in woman customer paths. Man customers usually spent their time in less visited places. For example, in Figure 4.7(c), man customers mainly visit Clothing and then leave the shopping mall, but they mostly spent their time in Catering and Electronics.

4.5 Discussion and Conclusion

In this study, we present a process mining implementation to discover man and woman customer paths in a shopping mall. This methodology has been tested using real three-month customer data belonging to from December 2017 to February 2018. The real data was collected using Bluetooth technology with iBeacon devices. Bluetooth is a cost effective tracking technology that provides unbiased and uninfluenced observations. We used a process mining algorithm that could cope with an infrequent and incomplete event log. PALIA Suite can consider infrequent behavior with statistical information about the activities frequencies and state changes. According to the results of the study, the implementation of process mining with ILS systems enables an understandable way.

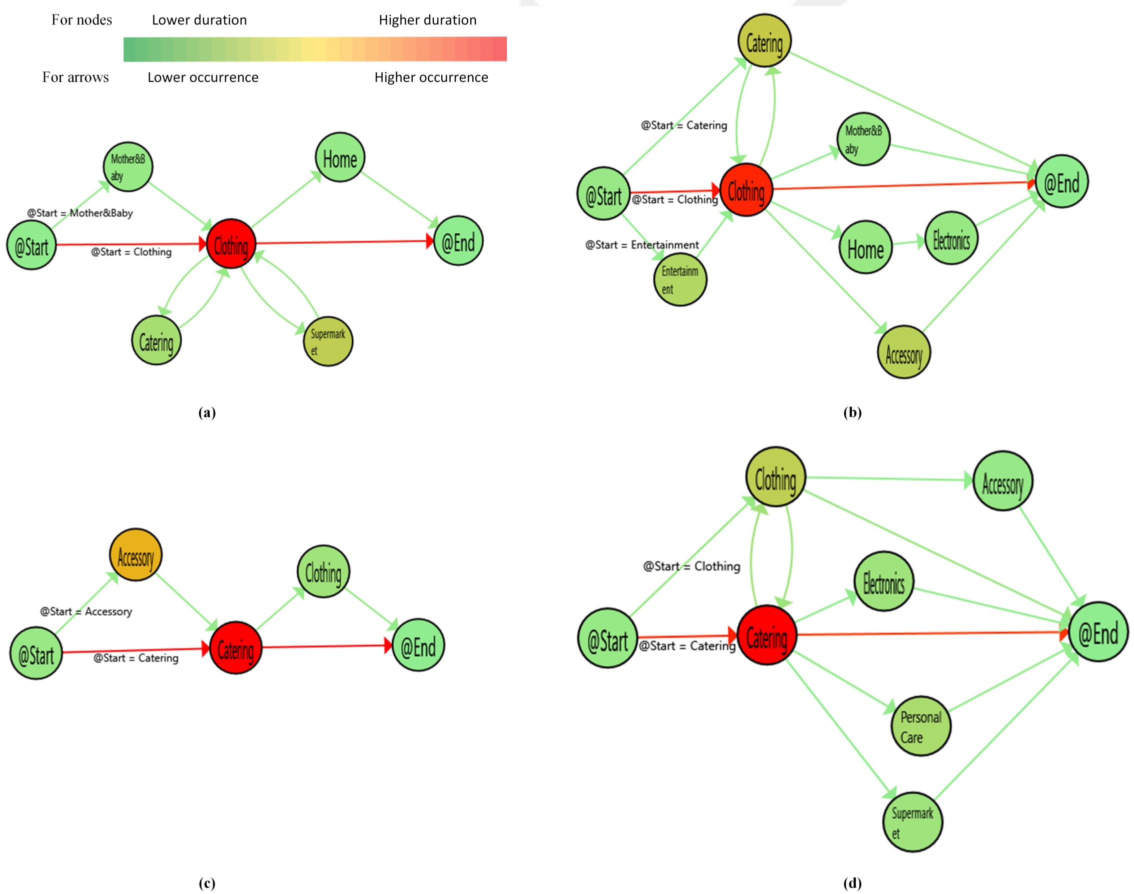

Figure 4.5 : Customers paths in December. (a): Male paths-1st cluster (b): Female paths-1st cluster (c): Male paths-2nd cluster (d): Female paths-2nd cluster.

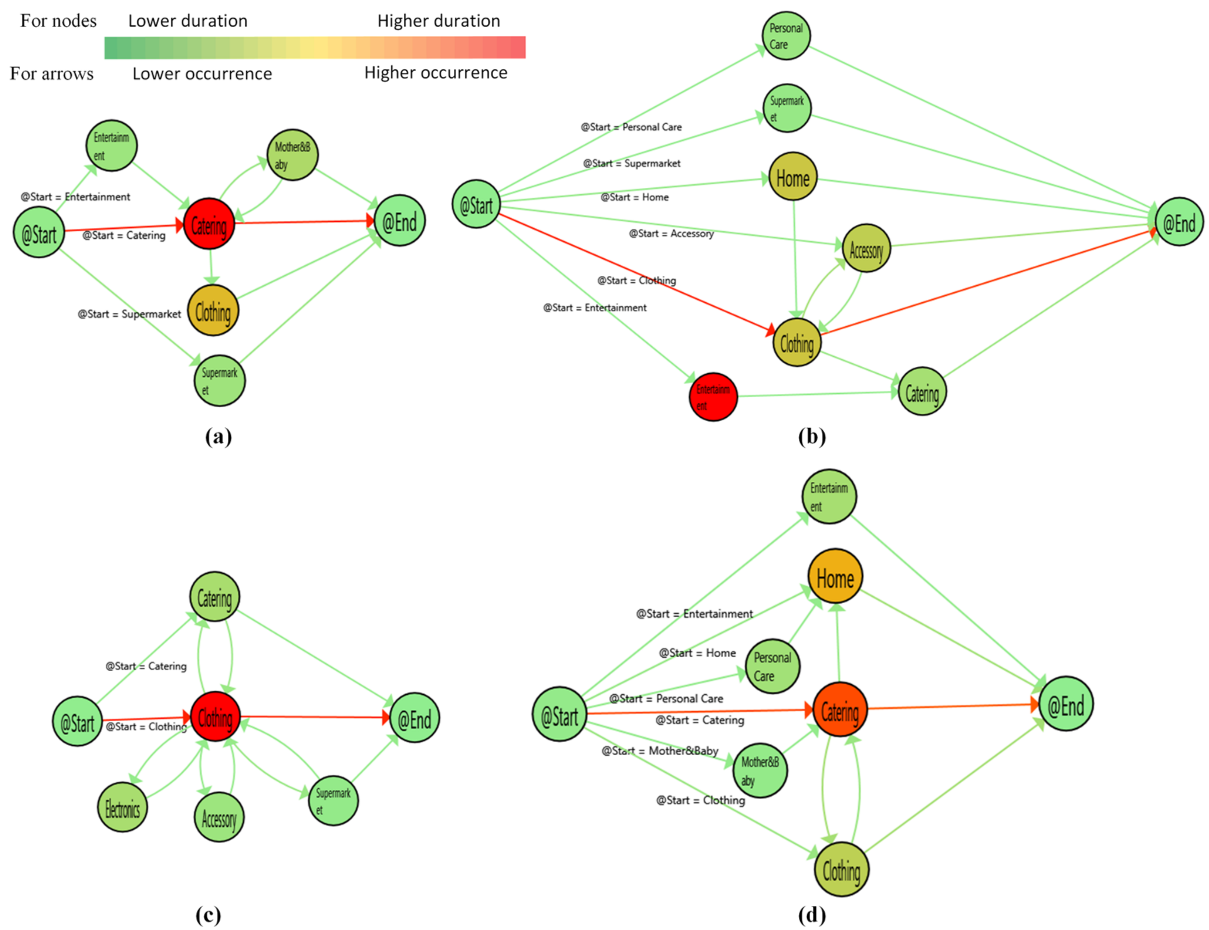

Figure 4.6 : Customers paths in January. (a): Male paths-1st cluster (b): Female paths-1st cluster (c): Male paths-2nd cluster (d): Female paths-2nd cluster.

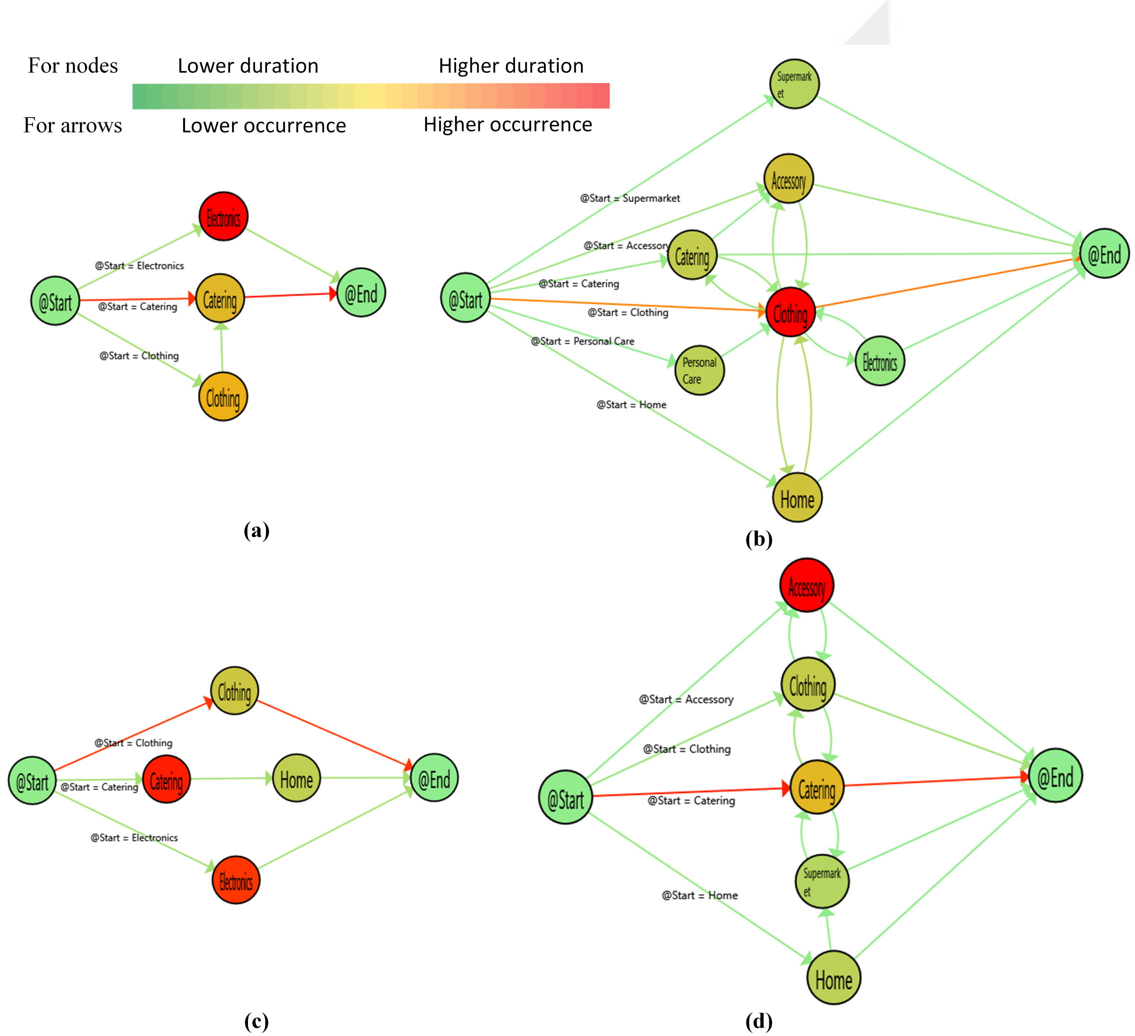

Figure 4.7 : Customers paths in February.(a): Male paths-1st cluster (b): Female paths-1st cluster (c): Male paths-2nd cluster (d): Female paths-2nd cluster.

We evaluate results regarding basic statistics and process mining. While basic statistics give some data –centric results, process mining is based on process oriented. We investigate 642 customer paths, which is 165 man and 477 woman customers. Since some customers visit the shopping mall more than one time, we create 1293 customer paths. The number of visited store groups is 2749 in this study. More customers visit the shopping mall in December, but more stores are visited in January. February is the weakest month with respect to the number of customers and visited stores. 9th December and 15th February are extremum dates. While 25 customers visited 59 store groups on 9th December, 7 customers visited 10 store groups on 15th February.

In December, Catering and Clothing are the most popular store groups. They are most visited and have the highest duration store groups. The numbers of customers who visit Catering then Clothing and vice versa they are 16 and 14 in man customer visits and 35 and 47 in woman customer visits, respectively. Man customers have a loop between Clothing – Catering and Clothing – Supermarket. On the other hand, woman customers have only Clothing – Catering loop. That may mean that woman customers are more decisive than man customers. We can also conclude that since man customers visit fewer store groups, they visit stores an unplanned way. Man customers visit fewer store groups than woman customers. Although Accessory is a common store group, woman customers spend less time. Moreover, Entertainment and Personal Care have significance concerning the behavioral difference. Man customers do not visit any Entertainment and Personal Care stores in mostly followed paths.

In January, Catering and Clothing are the most popular store groups for man customers like in December. On the other hand, the duration of Entertainment has increased for woman customers. The reason may be that some of the highly anticipated movies came out. Although woman customers in Figure 4.7(b) mostly visit Clothing, they spend more time in Entertainment. In many cases, customers visiting Entertainment visit a Catering store to eat or drink. The loops in man customer paths are still higher than woman customer paths like in December. When we compare December and January, man customers have again visited Catering and Clothing. Besides, Electronics and Mother&Baby have more duration in man customer visits. Home is another difference from December for woman customers in January. The number of visiting Home is nearly similar, but the duration in Home is four times larger than the duration in December.

In February, man customers prefer to visit Electronic store groups instead of Clothing. Catering is still one of the mostly visited locations. It is interesting that there is an inverse proportion between visited stores and duration in man paths. For example, man customers mostly visited Catering in Figure 4.7(a) and Clothing in Figure 4.7(c) but spent more time in another store groups, Electronics in Figure 4.7(a) and Catering and Electronics in Figure 4.7(b). On the contrary to the other two months, woman customers have higher loops. Woman customers visit Accessory store groups instead of Catering in this month. This may be because of Valentine's Day. However, man customers do not become interested in giving a gift on that special day.

In a general manner, it is interesting that customers who spend more time in Catering spend less time in Clothing and vice versa. Figures 4.5, 4.6 and 4.7 support the idea. Although man and woman customers have some similar traces, the duration is different. Catering has mainly the highest duration for man and woman customers. However, some store groups such as Entertainment and Accessory have become the critical points for behavioral changes in some cases.

We create a table that shows the number of visits and duration details in Table 4.7 to compare man and woman customers behaviors. In Table 4.7, Number_Occurrence shows the total number of visits, Duration_Total refers to total duration in each store groups, Duration_Average_By_Case presents the average duration with respect to the number of cases, and Duration_Average gives the average duration for each store groups. Duration_Average_By_Case and Duration_Average are different because one store group may be visited more than one time in any case by the same customer. For example, man customers visited Personal Care 8 times in December and spent 1 hour and 24 minutes. Since they visited Personal Care stores only one time in each case, Duration_Average_By_Case and Duration_Average are equal to 10 minutes and 29 seconds. On the other hand, woman customers visited Personal Care 33 times in December and spent 9 hours and 21 minutes. Since they visited Personal Care stores more than one time at least one case, Duration_Average_By_Case is 17 minutes and 31 seconds and Duration_Average is 17 minutes.

Table 4.6 : All comparison data.

	NUMBER_OCCURRENCES						DURATION_TOTAL					
	December		January		February		December		January		February	
NODES	Man	Woman	Man	Woman	Man	Woman	Man	Woman	Man	Woman	Man	Woman
Accessory	14	62	4	66	4	81	02:41:59	1.03:05:59	00:25:59	1.20:10:59	02:42:59	3.01:53:59
Catering	79	200	53	123	26	154	2.03:18:59	4.12:33:59	1.15:48:59	3.07:43:59	10:27:59	3.16:57:59
Clothing	93	223	60	189	17	242	9.18:33:59	4.00:46:59	8.19:34:59	5.14:16:59	05:10:59	17.18:42:59
Electronics	23	16	6	217	10	12	12:39:59	05:22:59	02:22:59	4.17:05:59	07:54:59	02:26:59
Entertainment	6	20	4	19	1	10	02:08:59	14:21:59	01:47:59	09:15:59	00:20:59	01:40:59
Home	24	77	10	252	6	76	06:51:59	1.04:35:59	02:44:59	5.11:04:59	06:54:59	2.01:46:59
Mother&Baby	10	24	6	27	0	31	02:30:59	09:08:59	02:09:59	11:00:59	00:00:00	2.02:08:59
Personal Care	8	33	0	22	1	13	01:23:59	09:20:59	00:00:00	05:37:59	01:03:59	06:32:59
Supermarket	19	26	7	23	5	15	05:44:59	11:26:59	02:28:59	07:25:59	00:54:59	05:15:59

	DURATION_AVERAGE_BY_CASE						DURATION_AVERAGE					
	December		January		February		December		January		February	
NODES	Man	Woman	Man	Woman	Man	Woman	Man	Woman	Man	Woman	Man	Woman
Accessory	00:13:29	00:29:33	00:06:29	00:47:20	00:40:44	01:04:15	00:11:34	00:26:13	00:06:29	00:40:09	00:40:44	00:54:44
Catering	00:44:37	00:35:24	00:46:50	00:42:20	00:24:09	00:37:19	00:38:58	00:32:34	00:45:04	00:38:53	00:24:09	00:34:39
Clothing	02:53:45	00:30:43	03:59:31	00:48:32	00:19:26	02:19:08	02:31:19	00:26:02	03:31:34	00:42:37	00:18:17	01:45:47
Electronics	00:34:32	00:20:11	00:23:49	02:41:34	00:47:29	00:12:14	00:33:02	00:20:11	00:23:49	00:31:16	00:47:29	00:12:14
Entertainment	00:25:47	00:45:22	00:26:59	00:30:53	00:20:59	00:10:05	00:21:29	00:43:05	00:26:59	00:29:15	00:20:59	00:10:05
Home	00:19:37	00:25:14	00:18:19	01:39:33	01:09:09	00:45:15	00:17:09	00:22:17	00:16:29	00:31:12	01:09:09	00:39:18
Mother&Baby	00:15:05	00:22:52	00:25:59	00:25:25	00:00:00	02:10:49	00:15:05	00:22:52	00:21:39	00:24:28	00:00:00	01:37:03
Personal Care	00:10:29	00:17:31	00:00:00	00:15:21	01:03:59	00:30:13	00:10:29	00:16:59	00:00:00	00:15:21	01:03:59	00:30:13
Supermarket	00:18:09	00:26:25	00:21:17	00:19:23	00:13:44	00:21:03	00:18:09	00:26:25	00:21:17	00:19:23	00:10:59	00:21:03

Table 4.7 : Probability matrix of transitions of male customers. $(x10^{-3})$.

Store Group	Accessory			Catering			Clothing			Electronics			Entertainment			Home			Mother&Baby			Personal Care			Supermarket		
	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F
Accessory				6.711	0.000	8.065	8.949	3.774	8.065	0.000	0.000	0.000	0.000	0.000	0.000	2.237	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.774	0.000
Catering	4.474	0.000	0.000				31.320	15.094	8.065	4.474	0.000	0.000	2.237	0.000	0.000	8.949	3.774	8.065	0.000	11.321	0.000	2.237	0.000	0.000	6.711	0.000	0.000
Clothing	4.474	3.774	8.065	35.794	15.094	24.194				6.711	3.774	8.065	0.000	3.774	0.000	6.711	0.000	0.000	2.237	3.774	0.000	2.237	0.000	0.000	11.186	7.547	0.000
Electronics	2.237	0.000	8.065	8.949	3.774	16.129	0.000	7.547	0.000				0.000	0.000	0.000	0.000	3.774	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Entertainment	2.237	0.000	0.000	0.000	3.774	0.000	0.000	0.000	8.065	0.000	0.000	0.000				0.000	0.000	0.000	4.474	3.774	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Home	4.474	3.774	0.000	4.474	0.000	0.000	11.186	0.000	0.000	2.237	7.547	0.000	0.000	0.000	0.000				0.000	3.774	0.000	2.237	0.000	0.000	6.711	0.000	8.065
Mother&Baby	0.000	0.000	0.000	2.237	3.774	0.000	2.237	3.774	0.000	4.474	0.000	0.000	0.000	0.000	0.000	2.237	0.000	0.000				0.000	0.000	0.000	0.000	0.000	0.000
Personal Care	0.000	0.000	0.000	2.237	0.000	0.000	4.474	0.000	8.065	0.000	0.000	0.000	2.237	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				2.237	0.000	0.000
Supermarket	0.000	0.000	0.000	2.237	0.000	0.000	6.711	7.547	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.237	3.774	8.065	4.474	0.000	0.000	0.000	0.000	0.000			

Table 4.8 : Probability matrix of transitions of female customers. $(x10^{-3})$.

Store Group	Accessory			Catering			Clothing			Electronics			Entertainment			Home			Mother&Baby			Personal Care			Supermarket		
	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F	D	J	F
Accessory				9.443	1.581	7.165	21.719	22.925	30.706	0.944	0.000	0.000	0.000	0.791	0.000	2.833	4.743	2.047	0.000	0.000	2.047	0.000	0.000	0.000	0.000	0.000	0.000
Catering	11.331	3.953	5.118				33.050	13.439	40.942	2.833	1.581	2.047	0.944	0.000	2.047	7.554	10.277	4.094	1.889	8.696	10.235	3.777	2.372	2.047	4.721	2.372	1.024
Clothing	17.941	18.972	24.565	44.381	26.877	50.154				2.833	1.581	4.094	4.721	0.000	1.024	9.443	4.743	23.541	4.721	2.372	5.118	7.554	3.953	3.071	4.721	3.162	3.071
Electronics	0.000	0.000	0.000	2.833	3.953	0.000	0.000	3.162	3.071				0.000	0.000	0.000	2.833	139.921	1.024	0.944	1.581	0.000	0.944	0.791	0.000	0.000	0.791	0.000
Entertainment	0.000	0.791	0.000	5.666	0.791	2.047	3.777	3.162	0.000	0.000	0.000	0.000				0.944	0.791	0.000	0.000	0.000	0.000	1.889	1.581	1.024	0.000	0.000	1.024
Home	4.721	3.953	1.024	7.554	5.534	5.118	13.220	5.534	33.777	0.000	146.245	0.000	0.944	0.791	0.000				0.944	0.000	1.024	1.889	1.581	0.000	5.666	3.162	3.071
Mother&Baby	0.000	0.791	2.047	2.833	3.953	4.094	4.721	3.162	5.118	0.000	0.791	1.024	0.000	0.000	0.000	0.944	2.372	1.024				1.889	0.000	0.000	0.000	0.000	0.000
Personal Care	0.000	0.000	0.000	4.721	0.791	4.094	2.833	1.581	1.024	0.944	1.581	0.000	1.889	1.581	0.000	2.833	0.791	0.000	1.889	0.000	0.000				0.944	0.000	0.000
Supermarket	0.000	0.000	0.000	0.944	3.162	2.047	4.721	1.581	1.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.581	0.000	0.944	0.000	0.000	0.944	0.791	0.000			

When we consider all paths for three months, Table 4.7 and Table 4.8 present the probability of all transitions among store groups. We realized exciting results from the matrix. All woman customers visit a Catering and Clothing store each month. Home is visited by all woman customers in December and January. Man customers visit Entertainment after either Catering or Personal Care in December, after Clothing in January. Although Catering and Clothing are two similar store groups for man and woman customers, the difference is that the possibility of visiting Clothing after Catering and Catering after Clothing is much higher in woman visits than man visits. Man customers visit Catering after Clothing by a rate of 0.031 and Clothing after Catering by a rate of 0,036 during December that has the highest transition probabilities. In the same month, woman customers have rates of 0.033 and 0.044 for transition probability from Catering to Clothing and Clothing to Catering, respectively. In January paths, man customers visit Personal Care as only the first visit point. They have no transition from other store groups to Personal Care. On the other hand, woman customers always visit Personal Care from other store groups except for Accessory and Mother&Baby. Entertainment is a critical store group in January, especially for woman customers. Woman customers visit Entertainment after Accessory, Home and Personal Care by a probability of 0.001, 0.001 and 0.002. After visiting Entertainment, they mainly visit either Clothing by a rate of 0.003 or return Personal Care by a rate of 0.002. However, man customers visit Entertainment after only one store group, Clothing, by a rate 0.004. Then, they continue to visit either Catering or Mother&Baby with equal transition probability, 0.004.

4.6 Future Research and Limitations

We consider all customer data in primary screening. Then, we focus on man and woman customer data to assess their behavior in secondary screening. As a future work, personalized guidance may be developed for customers. Customer-specific suggestions or future state prediction may be considered. Also, a process mining framework may be designed as in the study of Yang and Hwang [153]. They used process mining and introduced a framework to uncover fraud and abuse in healthcare. One of the advantages of process mining is to detect bottlenecks. Crowded in the shopping mall can be thought of as a bottleneck. Then, crowded store groups could be investigated. We have so much data in our corpus that we do not know the gender. We define genders whether customers go to the Toilette Man or Toilette Woman. For further research, we will predict the gender of customers from the follows paths. As a training data, we will use the known data that we use in this study.

Because of the technical properties of beacon software, only three localization data are collected. We cannot manipulate. We assume the location where the customer visits is shown by the nearest dongle. It is one of our assumptions. The study may be expanded by considering medium and far location data.

Use of Bluetooth for tracking purposes is an applicable approach, but there are some limitations. For example, customers must allow data sharing with the Bluetooth devices because the operating system often limits the access to customer's mobile. At the same time, although Bluetooth is a useful tracking technology, it is required to pre–install preparation.

5. DISCUSSION AND CONCLUSIONS

Understanding the needs of customers is the most basic requirement of today's companies. As many previous studies [22, 23] have supported, meeting customers' needs and offering personalized recommendations/campaigns is what customers expect. It is easier for companies operating online. However, the biggest challenge for physical stores is to build a system that will integrate online and offline channels. The first way to achieve this is to collect enough data to recognize customers. There are many ways of collecting data with developing the technology. Radio frequency identification (RFID), wireless connection (WiFi - wireless fidelity), Bluetooth are the most preferred ones. Apart from these, classical methods (observation, questionnaire, interview) and wireless communication technology (infrared, ZigBee and computer vision technologies such as camera and motion sensor) are the technology alternatives that can be used indoors to collect data. Since the selection of the most appropriate data collection technology for customer behavior analysis is by the multi-criteria decision making (MCDM) problem, this study was first focused on this issue. Fuzzy decision-making and the selection of the most suitable technology according to expert opinions were studied. As a result of the first study, Bluetooth was determined as the most suitable data collection technology. A shopping mall in Istanbul was equipped with iBeacon devices that work with Bluetooth technology. Thus, real data will be obtained without the human factor (without manual data entry), and the actual results will be analyzed by analyzing these data.

The second basic study is about finding the most frequently visited stores to facilitate customers' understanding. The discovery of the stores visited together provides many benefits for understanding the customers. We assume interesting rules include low support and high confidence values. The results can be used to propose personalized suggestion systems or to manage crowds in the shopping mall.

The final study is the generalization of the second study. In this study, the stores are divided into groups according to the products they sell or the services they provide. Behavior analysis was conducted by taking into consideration the routes of the customers between the store groups determined within the shopping mall. Process mining method has been used to create routes. One of the original aspects of the study is that the process mining method is used to extract customer routes in indoor locations. Also, the routes have been made more suitable for interpreting routes by quality threshold clustering. Path clustering is a part of the PALIA Suite algorithm. As a future work, different clustering algorithms can be developed for routes. Thus, it is possible to achieve better results. The newly developed hesitant fuzzy MCDM method is also extremely open to development. Any fuzzy decision-making method can be applied and the results comparable. The study shows there are differences and similarities among male and female customer behaviors in the shopping mall. The details are discussed in the third research. Because the number of male and female customers are very low, the results may have bias. To avoid this limitation, we predicted genders for gender-unknown customer. We used Levenshtein based fuzzy kNN method for gender prediction from indoor customer paths. It is under review in Expert Systems with Applications. After publication, we will use all three-month data to create customer paths.

REFERENCES

[1] **Hilbert, M. and López, P.** (2011). The world's technological capacity to store, communicate, and compute information, *Science*, *332*(6025), 60–65.

[2] **Manyika, J.**, **Chui, M.**, **Brown, B.**, **Bughin, J.**, **Dobbs, R.**, **Roxburgh, C. and Byers, A.H.** (2011). Big data: The next frontier for innovation, competition, and productivity.

[3] **Turner, V.**, **Gantz, J.F.**, **Reinsel, D. and Minton, S.** (2014). The digital universe of opportunities: Rich data and the increasing value of the internet of things, *IDC Analyze the Future*, *16*.

[4] **Yaqoob, I.**, **Hashem, I.A.T.**, **Gani, A.**, **Mokhtar, S.**, **Ahmed, E.**, **Anuar, N.B. and Vasilakos, A.V.** (2016). Big data: From beginning to future, *International Journal of Information Management*, *36*(6), 1231–1247.

[5] **Alharthi, A.**, **Krotov, V. and Bowman, M.** (2017). Addressing barriers to big data, *Business Horizons*, *60*(3), 285–292.

[6] **Van Der Aalst, W.** (2016). *Data Science in Action*, Springer, Berlin, Heidelberg.

[7] **Mou, S.**, **Robb, D.J. and DeHoratius, N.** (2018). Retail store operations: Literature review and research directions, *European Journal of Operational Research*, *265*(2), 399–422.

[8] **Cushman & Wakefield Inc**, (2012), A global perspective on the shopping center industry, Birmingham, UK.

[9] **Erkip, F. and Ozuduru, B.H.** (2015). Retail development in Turkey: An account after two decades of shopping malls in the urban scene, *Progress in Planning*, *102*, 1–33.

[10] **Andajani, E.** (2015). Understanding customer experience management in retailing, *Procedia-Social and Behavioral Sciences*, *211*, 629–633.

[11] **van der Aalst, W.M.P.** (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer.

[12] **van der Aalst, W.M.**, **van Dongen, B.F.**, **Herbst, J.**, **Maruster, L.**, **Schimm, G. and Weijters, A.J.** (2003). Workflow mining: A survey of issues and approaches, *Data & knowledge engineering*, *47*(2), 237–267.

[13] **Hwang, I. and Jang, Y.J.** (2017). Process mining to discover shoppers' pathways at a fashion retail store using a WiFi-base indoor positioning system, *IEEE Transactions on Automation Science and Engineering*, *14*(4), 1786–1792.

[14] **Dogan, O. and Oztaysi, B.** (2018). In-store behavioral analytics technology selection using fuzzy decision making, *Journal of Enterprise Information Management*, *31*(4), 612–630.

[15] **Celikkan, U.**, **Somun, G.**, **Kutuk, U.**, **Gamzeli, I.**, **Cinar, E.D. and Atici, I.** (2011). *Capturing supermarket shopper behavior using SmartBasket*, Springer.

[16] **Merad, D.**, **Aziz, K.E.**, **Iguernaissi, R.**, **Fertil, B. and Drap, P.** (2016). Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis, *Pattern Recognition Letters*, *81*, 11–20.

[17] **Infosys Corp.**, (2013), Consumers worldwide will allow access to personal data, United Kingtom.

[18] **Shave, L.**, (2016), Driving customer engagement inside and out of the store through personalized digital experiences.

[19] **Walker Corp.**, (2013), Customers 2020: The Future of B-to-B Customer Experience.

[20] **RetailNext Corp.**, (2017), Retail's Main Event: Brick & Mortar vs. Online.

[21] **Wu, Y.k.**, **Wang, H.C.**, **Chang, L.C. and Chou, S.C.** (2015). Customer's flow analysis in physical retail store, *Procedia Manufacturing*, *3*, 3506–3513.

[22] **Hurjui, C.**, **Graur, A. and Turcu, C.** (2008). Monitoring the shopping activities from the supermarkets based on the intelligent basket by using the RFID technology, *Elektronika ir Elektrotechnika*, *83*(3), 7–10.

[23] **Oosterlinck, D.**, **Benoit, D.F.**, **Baecke, P. and Van de Weghe, N.** (2017). Bluetooth tracking of humans in an indoor environment: An application to shopping mall visits, *Applied geography*, *78*, 55–65.

[24] **Huang, C.N. and Chan, C.T.** (2011). ZigBee-based indoor location system by k-nearest neighbor algorithm with weighted RSSI, *Procedia Computer Science*, *5*, 58–65.

[25] **Pan, M.S.**, **Liu, P.L. and Lin, Y.P.** (2014). Event data collection in ZigBee tree-based wireless sensor networks, *Computer Networks*, *73*, 142–153.

[26] **Liu, J.**, **Zhang, G.**, **Liu, Y.**, **Tian, L. and Chen, Y.Q.** (2015). An ultra-fast human detection method for color-depth camera, *Journal of Visual Communication and Image Representation*, *31*, 177–185.

[27] **López, Y.Á.**, **de Cos Gómez, M.E. and Andrés, F.L.H.** (2017). A received signal strength RFID-based indoor location system, *Sensors and Actuators A: Physical*, *255*, 118–133.

[28] **Jones, P.**, **Clarke-Hill, C.**, **Comfort, D.**, **Hillier, D. and Shears, P.** (2004). Radio frequency identification in retailing and privacy and public policy issues, *Management Research News*, *27*(8/9), 46–56.

[29] **Mohammadzadeh, F.F.**, **Liu, S.**, **Bond, K.A. and Nam, C.S.** (2015). Feasibility of a wearable, sensor-based motion tracking system, *Procedia Manufacturing*, *3*, 192–199.

[30] **Seol, S.**, **Lee, E.K. and Kim, W.** (2017). Indoor mobile object tracking using RFID, *Future Generation Computer Systems*, *76*, 443–451.

[31] **Yim, J.**, **Jeong, S.**, **Gwon, K. and Joo, J.** (2010). Improvement of Kalman filters for WLAN based indoor tracking, *Expert Systems with Applications*, *37*(1), 426–433.

[32] **Dzeng, R.J.**, **Lin, C.W. and Hsiao, F.Y.** (2014). Application of RFID tracking to the optimization of function-space assignment in buildings, *Automation in Construction*, *40*, 68–83.

[33] **Asha, C. and Narasimhadhan, A.** (2017). Robust infrared target tracking using discriminative and generative approaches, *Infrared Physics & Technology*, *85*, 114–127.

[34] **Bihler, P.**, **Imhoff, P. and Cremers, A.B.** (2011). SmartGuide–A smartphone museum guide with ultrasound control, *Procedia Computer Science*, *5*, 586–592.

[35] **Madhavapeddy, A.**, **Scott, D. and Sharp, R.** (2003). Context-Aware Computing with Sound, *UbiComp 2003: Ubiquitous Computing*, Springer Berlin Heidelberg, pp.315–332.

[36] **Saaty, T.L.** (1980). *The analytic hierarchy process : planning, priority setting, resource allocation*, McGraw-Hill International Book Co., New York; London.

[37] **Torra, V.** (2014). Hesitant fuzzy sets: State of the Art and Future Directions, *International Journal of Intelligent Systems*, *29*(6), 495–524.

[38] **Rodriguez, R.M.**, **Martinez, L. and Herrera, F.** (2012). Hesitant Fuzzy Linguistic Term Sets for Decision Making, *IEEE Transactions on Fuzzy Systems*, *20*(1), 109–119.

[39] **Onar, S.Ç.**, **Büyüközkan, G.**, **Öztayşi, B. and Kahraman, C.** (2016). A new hesitant fuzzy QFD approach: An application to computer workstation selection, *Applied Soft Computing*, *46*, 1–16.

[40] **Oztaysi, B.**, **Onar, S.C.**, **Bolturk, E. and Kahraman, C.** (2015). Hesitant fuzzy analytic hierarchy process, *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE.

[41] **Kirchberg, V. and Tröndle, M.** (2012). Experiencing Exhibitions: A Review of Studies on Visitor Experiences in Museums, *Curator: The Museum Journal*, *55*(4), 435–452.

[42] **Fernandez-Llatas, C.**, **Lizondo, A.**, **Monton, E.**, **Benedi, J.M. and Traver, V.** (2015). Process mining methodology for health process tracking using real-time indoor location systems, *Sensors*, *15*(12), 29821–29840.

[43] **Yuanfeng, D.**, **Dongkai, Y.**, **Huilin, Y. and Chundi, X.** (2016). Flexible indoor localization and tracking system based on mobile phone, *Journal of Network and Computer Applications*, *69*, 107–116.

[44] **Yoshimura, Y.**, **Sobolevsky, S.**, **Ratti, C.**, **Girardin, F.**, **Carrascal, J.P.**, **Blat, J. and Sinatra, R.** (2014). An analysis of visitors' behavior in the Louvre Museum: A study using Bluetooth data, *Environment and Planning B: Planning and Design*, *41*(6), 1113–1131.

[45] **Park, J.**, **Chen, J. and Cho, Y.K.** (2017). Self-corrective knowledge-based hybrid tracking system using BIM and multimodal sensors, *Advanced Engineering Informatics*, *32*, 126–138.

[46] **Ran, L.**, **Helal, S. and Moore, S.** (2004). Drishti: An integrated indoor/outdoor blind navigation system and service, *Proceedings - Second IEEE Annual Conference on Pervasive Computing and Communications, PerCom*, Orlando, FL, USA.

[47] **Lee, S.L.**, **Ibrahim, M.F. and Hsueh-Shan, C.** (2005). Shopping-centre attributes affecting male shopping behaviour, *Journal of Retail & Leisure Property*, *4*(4), 324–340.

[48] **Brosdahl, D.J. and Carpenter, J.M.** (2012). U.S. male generational cohorts: Retail format preferences, desired retail attributes, satisfaction and loyalty, *Journal of Retailing and Consumer Services*, *19*(6), 545–552.

[49] **Koklic, M.K.**, **Kukar-Kinney, M. and Vegelj, S.** (2017). An investigation of customer satisfaction with low-cost and full-service airline companies, *Journal of Business Research*, *80*, 188–196.

[50] **Bellini, S.**, **Cardinali, M.G. and Grandi, B.** (2017). A structural equation model of impulse buying behaviour in grocery retailing, *Journal of Retailing and Consumer Services*, *36*, 164–171.

[51] **Öztayşi, B.**, **Baysan, S. and Akpinar, F.** (2009). Radio frequency identification (RFID) in hospitality, *Technovation*, *29*(9), 618–624.

[52] **Clarke, R.H.**, **Twede, D.**, **Tazelaar, J.R. and Boyer, K.K.** (2006). Radio frequency identification (RFID) performance: the effect of tag orientation and package contents, *Packaging Technology and Science*, *19*(1), 45–54.

[53] **Roberts, C.** (2006). Radio frequency identification (RFID), *Computers & Security*, *25*(1), 18–26.

[54] **Cao, Q.**, **Jones, D.R. and Sheng, H.** (2014). Contained nomadic information environments: Technology, organization, and environment influences on adoption of hospital RFID patient tracking, *Information & Management*, *51*(2), 225–239.

[55] **Fernandez-Llatas, C.**, **Valdivieso, B.**, **Traver, V. and Benedi, J.M.** (2015). *Using Process Mining for Automatic Support of Clinical Pathways Design*, Springer, New York.

[56] **Ruotsalo, T.**, **Haav, K.**, **Stoyanov, A.**, **Roche, S.**, **Fani, E.**, **Deliai, R.**, **Mäkelä, E.**, **Kauppinen, T. and Hyvönen, E.** (2013). SMARTMUSEUM: A mobile recommender system for the Web of Data, *Journal of Web Semantics*, *20*, 50–67.

[57] **Curran, K.**, **Furey, E.**, **Lunney, T.**, **Santos, J.**, **Woods, D. and McCaughey, A.** (2011). An evaluation of indoor location determination technologies, *Journal of Location Based Services*, *5*(2), 61–78.

[58] **Fernández-Llatas, C.**, **Benedi, J.M.**, **García-Gómez, J.M. and Traver, V.** (2013). Process mining for individualized behavior modeling using wireless tracking in nursing homes, *Sensors*, *13*(11), 15434–15451.

[59] **Wu, C.**, **Yang, Z.**, **Liu, Y. and Xi, W.** (2013). WILL: Wireless Indoor Localization without Site Survey, *IEEE Transactions on Parallel and Distributed Systems*, *24*(4), 839–848.

[60] **Paciga, M. and Lutfiyya, H.** Herecast:an open infrastructure for locationbased services using WiFi, *WiMob2005), IEEE International Conference on Wireless And Mobile Computing, Networking And Communications, 2005.*, IEEE.

[61] **Villarubia, G.**, **Rubio, F.**, **Paz, J.F.D.**, **Bajo, J. and Zato, C.** (2013). *Applying Classifiers in Indoor Location System*, Springer International Publishing.

[62] **Kárník, J. and Streit, J.** (2016). Summary of available indoor location techniques, *IFAC-PapersOnLine*, *49*(25), 311–317.

[63] **Curran, K. and Furey, E.** (2008). Pinpointing users with location estimation techniques and Wi-Fi hotspot technology, *International Journal of Network Management*, *18*(5), 395–408.

[64] **Petre, A.C.**, **Chilipirea, C.**, **Baratchi, M.**, **Dobre, C. and van Steen, M.** (2017). *WiFi tracking of pedestrian behavior*, Elsevier.

[65] **Whitchurch, A.K.**, **Abraham, J.K. and Varadan, V.K.** (2007). Design and Development of a Wireless Remote Point-of-Care Patient Monitoring System, *2007 IEEE Region 5 Technical Conference*, IEEE.

[66] **Amato, F.**, **Chianese, A.**, **Mazzeo, A.**, **Moscato, V.**, **Picariello, A. and Picciall, F.** (2013). The Talking Museum Project, *Procedia Computer Science*, *21*, 114–121.

[67] **Castillo-Cara, M.**, **Huaranga-Junco, E.**, **Mondragón-Ruiz, G.**, **Salazar, A.**, **Barbosa, L.O. and Antúnez, E.A.** (2016). Ray: Smart Indoor/Outdoor Routes for the Blind Using Bluetooth 4.0 BLE, *Procedia Computer Science*, *83*, 690–694.

[68] **Delafontaine, M.**, **Versichele, M.**, **Neutens, T. and Van de Weghe, N.** (2012). Analysing spatiotemporal sequences in Bluetooth tracking data, *Applied Geography*, *34*, 659–668.

[69] **Stępień, J.**, **Kołodziej, J. and Machowski, W.** (2016). Mobile user tracking system with ZigBee, *Microprocessors and Microsystems*, *44*, 47–55.

[70] **Bu, F. and Chan, C.Y.** (2005). Pedestrian detection in transit bus application: sensing technologies and safety solutions, *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, IEEE.

[71] **Nguyen, D.T.**, **Li, W. and Ogunbona, P.O.** (2016). Human detection from images and videos: A survey, *Pattern Recognition*, *51*, 148–175.

[72] **Chen, S.**, **Fern, A. and Todorovic, S.** (2014). Multi-object tracking via constrained sequential labeling, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1130–1137.

[73] **Esterle, L.**, **Lewis, P.R.**, **Yao, X. and Rinner, B.** (2014). Socio-economic vision graph generation and handover in distributed smart camera networks, *ACM Transactions on Sensor Networks*, *10*(2), 1–24.

[74] **Shnayder, V.**, **rong Chen, B.**, **Lorincz, K.**, **Jones, T.R.F.F. and Welsh, M.** (2005). Sensor networks for medical care, *Proceedings of the 3rd international conference on Embedded networked sensor systems - SenSys 05*, ACM Press.

[75] **Mautz, R.** (2012). *Indoor Positioning Technologies*, (Ph.D. Thesis) ETH Zürich, Zurich.

[76] **Yucel, H.**, **Edizkan, R.**, **Ozkir, T. and Yazici, A.** (2012). Development of indoor positioning system with ultrasonic and infrared signals, *2012 International Symposium on Innovations in Intelligent Systems and Applications*, IEEE.

[77] **Svalastog, M.** (2007). *Indoor positioning-technologies, services and architectures*, (Ph.D. Thesis), University of Oslo, Norway.

[78] **Jovanovic, N.**, **Ozcelebi, T. and Lukkien, J.** (2014). Indoor user positioning using infrared LEDs and sensors, *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE.

[79] **Gonzalez, L.I.L.**, **Grobekathofert, U. and Amft, O.** (2014). Novel stochastic model for presence detection using ultrasound ranging sensors, *2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM)*, IEEE.

[80] **Frisby, J.**, **Smith, V.**, **Traub, S. and Patel, V.L.** (2017). Contextual computing: a Bluetooth based approach for tracking healthcare providers in the emergency room, *Journal of Biomedical Informatics*, *65*, 97–104.

[81] **Sadi-Nezhad, S. and Damghani, K.K.** (2010). Application of a fuzzy TOPSIS method base on modified preference ratio and fuzzy distance measurement in assessment of traffic police centers performance, *Applied Soft Computing*, *10*(4), 1028–1039.

[82] **Yager, R.** (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking, *IEEE Transactions on Systems, Man, and Cybernetics*, *18*(1), 183–190.

[83] **Filev, D. and Yager, R.R.** (1998). On the issue of obtaining OWA operator weights, *Fuzzy Sets and Systems*, *94*(2), 157–169.

[84] **van Laarhoven, P. and Pedrycz, W.** (1983). A fuzzy extension of Saaty's priority theory, *Fuzzy Sets and Systems*, *11*(1-3), 229–241.

[85] **Buckley, J.** (1985). Fuzzy hierarchical analysis, *Fuzzy Sets and Systems*, *17*(3), 233–247.

[86] **Cheng, C.H.** (1997). Evaluating naval tactical missile systems by fuzzy AHP based on the grade value of membership function, *European Journal of Operational Research*, *96*(2), 343–350.

[87] **Kahraman, C.**, **Öztayşi, B.**, **Sarı, İ.U. and Turanoğlu, E.** (2014). Fuzzy analytic hierarchy process with interval type-2 fuzzy sets, *Knowledge-Based Systems*, *59*, 48–57.

[88] **Otay, İ.**, **Oztaysi, B.**, **Onar, S.C. and Kahraman, C.** (2017). Multi-expert performance evaluation of healthcare institutions using an integrated intuitionistic fuzzy AHP&DEA methodology, *Knowledge-Based Systems*, *133*, 90–106.

[89] **Ilbahar, E.**, **Karaşan, A.**, **Cebi, S. and Kahraman, C.** (2018). A novel approach to risk assessment for occupational health and safety using Pythagorean fuzzy AHP & fuzzy inference system, *Safety Science*, *103*, 124–136.

[90] **Kumar, A.**, **Singh, P.**, **Kaur, A. and Kaur, P.** (2010). RM Approach for Ranking of Generalized Trapezoidal Fuzzy Numbers, *Fuzzy Information and Engineering*, *2*(1), 37–47.

[91] **Li, Y.M.**, **Lin, L.F. and Ho, C.C.** (2017). A social route recommender mechanism for store shopping support, *Decision Support Systems*, *94*, 97–108.

[92] **Das, G. and Varshneya, G.** (2017). Consumer emotions: Determinants and outcomes in a shopping mall, *Journal of Retailing and Consumer Services*, *38*, 177–185.

[93] **Kesari, B. and Atulkar, S.** (2016). Satisfaction of mall shoppers: A study on perceived utilitarian and hedonic shopping values, *Journal of Retailing and Consumer Services*, *31*, 22–31.

[94] **Yiu, C.Y. and Ng, H.C.** (2010). Buyers-to-shoppers ratio of shopping malls: A probit study in Hong Kong, *Journal of Retailing and Consumer Services*, *17*(5), 349–354.

[95] **Willems, K.**, **Smolders, A.**, **Brengman, M.**, **Luyten, K. and Schöning, J.** (2017). The path-to-purchase is paved with digital opportunities: An inventory of shopper-oriented retail technologies, *Technological Forecasting and Social Change*, *124*, 228–242.

[96] **V., J.L.C.**, **Zhao, Z.**, **Braun, T.**, **Li, Z. and Neto, A.** (2018). A real-time robust indoor tracking system in smartphones, *Computer Communications*, *117*, 104–115.

[97] **Fukuzaki, Y., Mochizuki, M., Murao, K. and Nishio, N.** (2015). Statistical analysis of actual number of pedestrians for Wi-Fi packet-based pedestrian flow sensing, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp 15*, ACM Press.

[98] **Yewatkar, A., Inamdar, F., Singh, R., Ayushya and Bandal, A.** (2016). Smart Cart with Automatic Billing, Product Information, Product Recommendation Using RFID & Zigbee with Anti-Theft, *Procedia Computer Science*, *79*, 793–800.

[99] **Arroyo, R., Yebes, J.J., Bergasa, L.M., Daza, I.G. and Almazán, J.** (2015). Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls, *Expert systems with Applications*, *42*(21), 7991–8005.

[100] **Yoshimura, Y., Krebs, A. and Ratti, C.** (2017). Noninvasive Bluetooth Monitoring of Visitors, *IEEE Pervasive Computing*, *16*(2), 26–34.

[101] **Salim, F., Williams, M., Sony, N., Pena, M.D., Petrov, Y., Saad, A.A. and Wu, B.** (2014). Visualization of wireless sensor networks using ZigBees Received Signal Strength Indicator (RSSI) for indoor localization and tracking, *2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*, IEEE.

[102] **Catarinucci, L., Colella, R., Mainetti, L., Patrono, L., Pieretti, S., Sergi, I. and Tarricone, L.** (2014). Smart RFID Antenna System for Indoor Tracking and Behavior Analysis of Small Animals in Colony Cages, *IEEE Sensors Journal*, *14*(4), 1198–1206.

[103] **Liu, H., Yang, J., Sidhom, S., Wang, Y., Chen, Y. and Ye, F.** (2014). Accurate WiFi Based Localization for Smartphones Using Peer Assistance, *IEEE Transactions on Mobile Computing*, *13*(10), 2199–2214.

[104] **Hernandez, N., Ocana, M., Alonso, J.M. and Kim, E.** (2014). WiFi-based indoor localization and tracking of a moving device, *2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS)*, IEEE.

[105] **Jaramillo, P. and Linnartz, J.P.** (2015). Hidden Markov Model for Improved Ultrasound-Based Presence Detection, *2015 IEEE International Conference on Computer and Information Technology Ubiquitous Computing and Communications Dependable, Autonomic and Secure Computing Pervasive Intelligence and Computing*, IEEE.

[106] **Feng, F., Cho, J., Pedrycz, W., Fujita, H. and Herawan, T.** (2016). Soft set based association rule mining, *Knowledge-Based Systems*, *111*, 268–282.

[107] **Agrawal, R., Imieliński, T. and Swami, A.** (1993). Mining association rules between sets of items in large databases, *ACM SIGMOD Record*, *22*(2), 207–216.

[108] **Tan, P.N.**, **Steinbach, M. and Kumar, V.** (2005). *Introduction to Data Mining*, Pearson.

[109] **Han, J.**, **Kamber, M. and Pei, J.** (2012). *Data Mining: Concepts and Techniques*, Elsevier.

[110] **Han, J.**, **Pei, J. and Yin, Y.** (2000). Mining Frequent Patterns Without Candidate Generation, *SIGMOD Recor*, *29*(2), 1–12.

[111] **Abedi, N.**, **Bhaskar, A.**, **Chung, E. and Miska, M.** (2015). Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses, *Transportation Research Part C: Emerging Technologies*, *60*, 124–141.

[112] **Mou, S.**, **Robb, D.J. and DeHoratius, N.** (2017). Retail store operations: Literature review and research directions, *European Journal of Operational Research*, *265*(2), 399–422.

[113] **Ou-Yang, C. and Winarjo, H.** (2011). Petri-net integration–An approach to support multi-agent process mining, *Expert systems with applications*, *38*(4), 4039–4051.

[114] **Partington, A.**, **Wynn, M.**, **Suriadi, S.**, **Ouyang, C. and Karnon, J.** (2015). Process mining for clinical processes: A comparative analysis of four Australian hospitals, *ACM Transactions on Management Information Systems (TMIS)*, *5*(4), 19.

[115] **Yoo, S.**, **Cho, M.**, **Kim, E.**, **Kim, S.**, **Sim, Y.**, **Yoo, D.**, **Hwang, H. and Song, M.** (2016). Assessment of hospital processes using a process mining technique: outpatient process analysis at a tertiary hospital, *International journal of medical informatics*, *88*, 34–43.

[116] **Funkner, A.A.**, **Yakovlev, A.N. and Kovalchuk, S.V.** (2017). Towards evolutionary discovery of typical clinical pathways in electronic health records, *Procedia Computer Science*, *119*, 234–244.

[117] **Jans, M.**, **Alles, M. and Vasarhelyi, M.** (2013). The case for process mining in auditing: Sources of value added and areas of application, *International Journal of Accounting Information Systems*, *14*(1), 1–20.

[118] **De Leoni, M.**, **van der Aalst, W.M. and Dees, M.** (2016). A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs, *Information Systems*, *56*, 235–257.

[119] **Rebuge, Á. and Ferreira, D.R.** (2012). Business process analysis in healthcare environments: A methodology based on process mining, *Information systems*, *37*(2), 99–116.

[120] **Popa, M.C.**, **Rothkrantz, L.J.**, **Shan, C.**, **Gritti, T. and Wiggers, P.** (2013). Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information, *Pattern Recognition Letters*, *34*(7), 809–819.

[121] **Kang, L. and Hansen, M.** (2017). Behavioral analysis of airline scheduled block time adjustment, *Transportation Research Part E: Logistics and Transportation Review*, *103*, 56–68.

[122] **Rovani, M.**, **Maggi, F.M.**, **de Leoni, M. and van der Aalst, W.M.** (2015). Declarative process mining in healthcare, *Expert Systems with Applications*, *42*(23), 9236–9251.

[123] **Mans, R.**, **Reijers, H.**, **van Genuchten, M. and Wismeijer, D.** (2012). Mining processes in dentistry, *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, ACM, pp.379–388.

[124] **van der Aalst, W.M.**, **Reijers, H.A.**, **Weijters, A.J.**, **van Dongen, B.F.**, **de Medeiros, A.A.**, **Song, M. and Verbeek, H.** (2007). Business process mining: An industrial application, *Information Systems*, *32*(5), 713–732.

[125] **Valle, A.M.**, **Santos, E.A. and Loures, E.R.** (2017). Applying process mining techniques in software process appraisals, *Information and software technology*, *87*, 19–31.

[126] **Rubin, V.A.**, **Mitsyuk, A.A.**, **Lomazova, I.A. and van der Aalst, W.M.** (2014). Process mining can be applied to software tool, *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, ACM, p. 57.

[127] **Gupta, M.**, **Sureka, A. and Padmanabhuni, S.** (2014). Process mining multiple repositories for software defect resolution from control and organizational perspective, *Proceedings of the 11th Working Conference on Mining Software Repositories*, ACM, pp.122–131.

[128] **Juhaňák, L.**, **Zounek, J. and Rohlíková, L.** (2017). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system, *Computers in Human Behavior*, *92*, 496–506.

[129] **Sedrakyan, G.**, **De Weerdt, J. and Snoeck, M.** (2016). Process-mining enabled feedback:"Tell me what I did wrong" vs."tell me how to do it right", *Computers in human behavior*, *57*, 352–376.

[130] **Schoor, C. and Bannert, M.** (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining, *Computers in Human Behavior*, *28*(4), 1321–1331.

[131] **Werner, M. and Gehrke, N.** (2015). Multilevel process mining for financial audits, *IEEE Transactions on Services Computing*, *8*(6), 820–832.

[132] **de Weerdt, J.**, **Schupp, A.**, **Vanderloock, A. and Baesens, B.** (2013). Process Mining for the multi-faceted analysis of business processes—A case study in a financial services organization, *Computers in Industry*, *64*(1), 57–67.

[133] **Herbert, L.**, **Hansen, Z.N.L.**, **Jacobsen, P. and Cunha, P.** (2014). Evolutionary optimization of production materials workflow processes, *Procedia CIRP*, *25*, 53–60.

[134] **van Eck, M.L.**, **Sidorova, N. and van der Aalst, W.M.** (2016). Enabling process mining on sensor data from smart products, *Research Challenges in Information Science (RCIS), 2016 IEEE Tenth International Conference on*, IEEE, pp.1–12.

[135] **Diamantini, C.**, **Genga, L.**, **Marozzo, F.**, **Potena, D. and Trunfio, P.** (2017). Discovering Mobility Patterns of Instagram Users through Process Mining Techniques, *Information Reuse and Integration (IRI), 2017 IEEE International Conference on*, IEEE, pp.485–492.

[136] **Yoshimura, Y.**, **Girardin, F.**, **Carrascal, J.P.**, **Ratti, C. and Blat, J.** (2012). *New tools for studying visitor behaviours in museums: A case study at the Louvre*, Springer-Verlag.

[137] **Yoshimura, Y.**, **Krebs, A. and Ratti, C.** (2016). An analysis of visitors' length of stay through noninvasive Bluetooth monitoring in the Louvre Museum, *ArXiv Preprint*.

[138] **Larson, J.S.**, **Bradlow, E.T. and Fader, P.S.** (2005). An exploratory look at supermarket shopping paths, *International Journal of research in Marketing*, *22*(4), 395–414.

[139] **Vukovic, M.**, **Lovrek, I. and Kraljevic, H.** (2012). *Discovering shoppers' journey in retail environment by using RFID.*, IOS Press.

[140] **Leemans, S.J.**, **Fahland, D. and van der Aalst, W.M.P.** (2014). Discovering block-structured process models from incomplete event logs, *International Conference on Applications and Theory of Petri Nets and Concurrency*, Springer, pp.91–110.

[141] **Buijs, J.C.A.M.** (2014). *Flexible evolutionary algorithms for mining structured process models*, (Ph.D. Thesis), Technische Universiteit Eindhoven.

[142] **Fernandez-Llatas, C.**, **Martinez-Millana, A.**, **Martinez-Romero, A.**, **Benedi, J.M. and Traver, V.** (2015). Diabetes care related process modelling using Process Miningtechniques. Lessons learned in the application of InteractivePattern Recognition: coping with the Spaghetti Effect, *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp.2127–2130.

[143] **Conca, T.**, **Saint-Pierre, C.**, **Herskovic, V.**, **Sepúlveda, M.**, **Capurro, D.**, **Prieto, F. and Fernandez-Llatas, C.** (2018). Multidisciplinary Collaboration in the Treatment of Patients With Type 2 Diabetes in Primary Care: Analysis Using Process Mining, *Journal of medical Internet research*, *20*(4).

[144] **Bahl, P. and Padmanabhan, V.N.** (2000). RADAR: An in-building RF-based user location and tracking system, *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE.*, volume 2, IEEE, pp.775–784.

[145] **Mans, R., van der Aalst, W. and Vanwersch, R.J.** (2015). *Process mining in healthcare:: Evaluating and exploiting operational healthcare processes*, Springer.

[146] **Weijters, A. and Ribeiro, J.** (2011). Flexible heuristics miner (FHM), *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on*, IEEE, pp.310–317.

[147] **de Medeiros, A.K.A., Weijters, A.J. and van der Aalst, W.M.** (2007). Genetic process mining: an experimental evaluation, *Data Mining and Knowledge Discovery*, *14*(2), 245–304.

[148] **Fernández-Llatas, C., Meneu, T., Benedi, J.M. and Traver, V.** (2010). Activity-based process mining for clinical pathways computer aided design, *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, IEEE, pp.6178–6181.

[149] **Fernandez-Llatas, C., Pileggi, S.F., Traver, V. and Benedi, J.M.** (2011). Timed parallel automaton: A mathematical tool for defining highly expressive formal workflows, *Modelling Symposium (AMS), 2011 Fifth Asia*, IEEE, pp.56–61.

[150] **Heyer, L.J., Kruglyak, S. and Yooseph, S.** (1999). Exploring expression data: identification and analysis of coexpressed genes, *Genome research*, *9*(11), 1106–1115.

[151] **Vidal, E., Rulot, H., Valiente, J.M. and Andreu, G.** (1993). Application of the error-correcting grammatical inference algorithm (ECGI) to planar shape recognition, *IEE Colloquium on Grammatical Inference: Theory, Applications and Alternatives*, pp.24/1–2410.

[152] **Oztaysi, B., Gokdere, U., Simsek, E.N. and Oner, C.S.** (2017). *A novel approach to segmentation using customer locations data and intelligent techniques*, IGI Global.

[153] **Yang, W.S. and Hwang, S.Y.** (2006). A process-mining framework for the detection of healthcare fraud and abuse, *Expert Systems with Applications*, *31*(1), 56–68.

CURRICULUM VITAE

Name Surname: ONUR DOĞAN

Place and Date of Birth: Fatsa/Ordu/Turkey 29.08.1987

E-Mail: onurdoganmail@gmail.com

EDUCATION:

- **B.Sc.:** 2010, Sakarya University, Faculty of Engineering, Department of Industrial Engineering
- **M.Sc.:** 2014, Istanbul Technical University, Management Faculty, Department of Industrial Engineering year, University, Faculty, Department
- **Visiting Researcher:** 2018, Instituto Universitario de Investigación de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2011-2019 Research Assistant at Department of Industrial Engineering, Istanbul Technical University.
- 2018-2019 Visitor Researcher at Instituto Universitario de Investigación de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones
- Avanzadas (ITACA), Universitat Politècnica de València with TUBITAK scholarship for 12 months
- 2010-2011 Research Assistant at Department of Industrial Engineering, Karabuk University

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Dogan, O.** Bayo-Montón, J.L., Fernandez-Llatas, C., Oztaysi, B. (2019). Analyzing of Gender Behaviors from Paths Using Process Mining: A Shopping Mall Application. *Sensors 19*(3), 557-577.
- **Dogan, O.**, Gurcan, O.F., Oztaysi, B. and Gokdere, U. (2019). Analysis of Frequent Visitor Patterns in a Shopping Mall" in *Industrial Engineering in the Big Data Era*, (p. 217-227), Springer.
- **Dogan, O.**, Oztaysi, B. (2018). In-store behavioral analytics technology selection using fuzzy decision making. *Journal of Enterprise Information Management, 31*(4), 612-630.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

Dogan, O.. (2018). Process Mining For Check-Up Process Analysis, *IIOABJ* , *29*(6), p. 56-61.

Kahraman, C., Oztaysi, B., Cevik Onar, S., **Dogan, O.** (2018). Intuitionistic Fuzzy Originated Type-2 Fuzzy AHP: An Application to Damless Hydroelectric Power Plants, *International Journal of the Analytic Hierarchy Process. 8*(2), 266-292.

Dogan, O., Gurcan, O.F. (2019). Applications of Big Data and Green IoT Enabling Technologies for Smart Cities" in *Handbook of Research on Big Data and the IoT*, IGI Global.

Dogan, O. (2018). Data Linkage Methods For Big Data Management in Industry 4.0 in *Optimizing Big Data Management and Industrial Systems With Intelligent Techniques*, (p. 108-127), IGI Global.

Dogan, O. (2017). Heuristic Approaches for Clustering Problems in *Applied Optimization Methodologies in Manufacturing Systems*, (p. 107-124), IGI Global.