ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

DATA-DRIVEN MODELLING OF COST ADVANTAGE STRATEGIES ENHANCING OPERATIONAL EFFICIENCY IN THE RETAIL SECTOR

Ph.D. THESIS

Çağatay ÖZDEMİR

Department of Industrial Engineering

Industrial Engineering Programme

MAY 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

DATA-DRIVEN MODELLING OF COST ADVANTAGE STRATEGIES ENHANCING OPERATIONAL EFFICIENCY IN THE RETAIL SECTOR

Ph.D. THESIS

Çağatay ÖZDEMİR
(507202101)

Department of Industrial Engineering

Industrial Engineering Programme

Thesis Advisor: Prof. Dr. Sezi Çevik ONAR

MAY 2025

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

PERAKENDE SEKTÖRÜNDE OPERASYONEL VERİMLİLİĞİ ARTTIRAN MALİYET AVANTAJI STRATEJİLERİNİN VERİ TEMELLİ MODELLENMESİ

DOKTORA TEZİ

Çağatay ÖZDEMİR
(507202101)

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Sezi Çevik ONAR

MAYIS 2025

Çağatay ÖZDEMİR, a Ph.D. student of İTU Graduate School student ID 507202101 successfully defended the thesis/dissertation entitled "DATA-DRIVEN MODELLING OF COST ADVANTAGE STRATEGIES ENHANCING OPERATIONAL EFFICIENCY IN THE RETAIL SECTOR", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Sezi Çevik ONAR** ....
İstanbul Technical University

Jury Members : **Prof. Dr. Başar ÖZTAYŞİ** ....
Istanbul Technical University

Assoc. Prof. Dr. Irem OTAY ....
Istanbul Bilgi University

Prof. Dr. Cengiz KAHRAMAN ....
Istanbul Technical University

Prof. Dr. Selçuk ÇEBİ ....
Yıldız Technical University

Date of Submission : 21 April 2025
Date of Defense : 30 May 2025

To my family,

FOREWORD

In the context of this thesis on data-based cost advantage models for retail, comprehensive research was conducted on data mining methodologies, customer management strategies, and transaction cost theories. The shopping mall visitor prediction study involved integrating demand forecasting with location analysis to develop a novel hybrid fuzzy prediction method incorporating environmental factors for improved accuracy. Similarly, research for online retail delivery optimization and product development decision-making frameworks required extensive exploration of machine learning techniques, fuzzy multi-criteria decision-making methods, and economic theories to create innovative analytical approaches that bridge theoretical concepts with practical applications.

I want to express my sincere gratitude to my advisor, Prof. Dr. Sezi Çevik ONAR, whose guidance and expertise were instrumental throughout this research journey, which began from a bachelor's degree to a Ph.D. degree. I would like to thank Prof. Dr. Başar ÖZTAYŞİ and Assoc. Prof. Dr. Irem OTAY, who were in my thesis jury, for their guidance and support throughout the thesis process.

I would also like to thank my beloved wife Zeynep Hazal AYSAN ÖZDEMİR and my beloved daughter İnci Ferda ÖZDEMİR, who have always been with me and supported me throughout this process, and my father Sait ÖZDEMİR and my mother Gülsüm ÖZDEMİR, who have made great efforts to help me reach this degree.

I specially want to express my gratitude and respect to Gazi Mustafa Kemal ATATÜRK and his comrades-in-arms, and to the REPUBLIC OF TÜRKİYE, which has given us scientific thought and the opportunity to conduct research.

May 2025

Çağatay ÖZDEMİR
(Industrial Engineer MSc)

TABLE OF CONTENTS

Page

ABBREVIATIONS

AI	: Artificial Intelligence
AHP	: Analytic Hierarchy Process
AUROC	: Area Under the Receiver Operating Characteristic
B2C	: Business-to-Consumer
CNN-US	: Condensed Nearest Neighbor Under Sampling
EDA	: Exploratory Data Analysis
FAHP	: Fuzzy Analytic Hierarchy Process
GBM	: Gradient Boosting Method
MCDM	: Multi-Criteria Decision Making
ML	: Machine Learning
MSE	: Mean Squared Error
OLS	: Ordinary Least Squares
PIS	: Positive Ideal Solution
ROC	: Receiver Operating Characteristic
ROI	: Return on Investment
SFNIS	: Spherical Fuzzy Negative Ideal Solution
SFPIS	: Spherical Fuzzy Positive Ideal Solution
SFS	: Spherical Fuzzy Sets
SGBM	: Stochastic Gradient Boosting Method
SHAP	: SHapley Additive exPlanations
SWAM	: Spherical Fuzzy Weighted Arithmetic Mean
SWGM	: Spherical Fuzzy Weighted Geometric Mean
TOPSIS	: Technique for Order of Preference by Similarity to Ideal Solution
VIF	: Variance Inflation Factor

SYMBOLS

$\mathbf{a_{ij}}$	: Elements of pairwise comparison matrix
$\boldsymbol{\mu}$	: Membership degree in spherical fuzzy sets
$\boldsymbol{\nu}$	: Non-membership degree in spherical fuzzy sets
$\boldsymbol{\pi}$	: Neutrality degree in spherical fuzzy sets
$\mathbf{S_i}$	: Fuzzy synthetic extent with respect to the i-th object
$\mathbf{R^2}$	: Coefficient of determination
$\mathbf{R^2_{adj}}$	: Adjusted coefficient of determination
$\mathbf{SS_{RES}}$	: Sum of squares of residuals
$\mathbf{SS_{Tot}}$	: Total sum of squares
$\boldsymbol{\Delta}$	: Delta (change in value)
$\tilde{\mathbf{A}}$	: Fuzzy comparison matrix
$\tilde{\boldsymbol{r}}_{\boldsymbol{ij}}$	: Normalized fuzzy values
$\widetilde{\boldsymbol{U}}_{\boldsymbol{i}}$	: Aggregated fuzzy weights
$\boldsymbol{D}(\boldsymbol{Xi}, \boldsymbol{X}*)$	: Distance between alternative and positive ideal solution
$\boldsymbol{D}(\boldsymbol{Xi}, \boldsymbol{X}-)$	: Distance between alternative and negative ideal solution

LIST OF TABLES

Page

LIST OF FIGURES

Page

DATA-DRIVEN MODELLING OF COST ADVANTAGE STRATEGIES ENHANCING OPERATIONAL EFFICIENCY IN THE RETAIL SECTOR

SUMMARY

This thesis explores developing and implementing data-based cost advantage models in the retail sector, covering both physical and online retail environments. The research addresses a significant gap in the literature by proposing an integrated approach that combines machine learning algorithms with fuzzy decision-making techniques to optimize retail operations and enhance cost efficiency.

The study consists of three interconnected research areas. The first focuses on predicting shopping mall visitor numbers through a novel hybrid fuzzy prediction method. This approach integrates machine learning algorithms with fuzzy multi-criteria decision-making techniques to analyze the impact of environmental factors such as weather conditions, financial market data, and internet trends on visitor patterns. Data from three shopping malls on the Anatolian side of Istanbul validated the model, demonstrating its high prediction accuracy and generalizability.

The second research area examines resource allocation decisions in product development processes through the lens of transaction cost economics. The study identifies and prioritizes cost factors that influence internal versus external resource allocation decisions by applying the Fuzzy Analytic Hierarchy Process to transaction cost theory. The findings reveal that infrastructure and development costs (44%) and uncertainty and time costs (31%) are the most significant factors affecting these decisions.

The third area addresses delivery optimization in online retail by developing a hybrid approach that combines machine learning with Spherical Fuzzy TOPSIS methodology. This innovative integration transforms SHAP (Shapley Additive exPlanations) values from machine learning models into Spherical Fuzzy Sets, enabling more effective decision-making under uncertainty. The approach was applied to evaluate different delivery strategies (multi-package versus immediate delivery), providing online retailers with a framework to balance operational efficiency and customer satisfaction.

The thesis contributes to both theory and practice by advancing the methodological integration of machine learning with fuzzy decision-making techniques, extending transaction cost theory to product development decision-making, and introducing a novel approach to modeling uncertainty in retail decision processes. The proposed frameworks offer retailers practical tools to optimize operations, reduce costs, and enhance competitiveness in an increasingly complex market environment.

PERAKENDE SEKTÖRÜNDE OPERASYONEL VERİMLİLİĞİ ARTTIRAN MALİYET AVANTAJI STRATEJİLERİNİN VERİ TEMELLİ MODELLENMESİ

ÖZET

Bu tez, perakende sektöründe veri tabanlı maliyet avantajı modellerinin geliştirilmesi ve uygulanmasını, hem fiziksel hem de internet üzerinden perakende ortamlarını kapsayacak şekilde incelemektedir. Perakende sektörü içerisinde bulundurduğu alt sektörler ile birlikte ekonominin en dinamik ve hızla dönüşen alanlarından biridir. Son yıllarda bu sektör yapısal olarak üç temel alt sektör grubuna ayrılmaktadır: fiziksel/yerinde (offline) perakende, internet üzerinden (online) perakende ve ikisinin de aynı anda kullanabildiği hibrit perakende modelleri. Dijital dönüşümün hızlanmasıyla birlikte, perakende sektöründe üretilen veri miktarı ve çeşitliliği önemli ölçüde artmış, bu verilerin analizi ve stratejik kararlarda kullanımı rekabet avantajı elde etmenin en önemli unsurlarından biri haline gelmiştir.

Günümüzde veri tabanlı modeller oluşturulurken karşılaşılan en önemli problemlerden bazıları, modelin kurulacağı doğru problemin seçilememesi, seçilen probleme uygun verilerin belirlenmesi ve toplanamaması, verilerin doğru metodolojilerle analiz edilememesi ve nihai olarak kurulan modelin doğru ürünleştirecek senaryoya dönüştürülememesidir. Bu zorluklar, veri tabanlı modellerin potansiyel faydalarının tam olarak realize edilememesine ve yapılan yatırımların beklenen getiriyi sağlayamamasına neden olmaktadır. Bu araştırma, makine öğrenmesi algoritmaları ile bulanık karar verme tekniklerini birleştiren entegre bir yaklaşım önererek, perakende operasyonlarını optimize etmek ve maliyet verimliliğini artırmak amacıyla literatürdeki önemli bir boşluğu doldurmaktadır.

Çalışma, birbiriyle bağlantılı üç araştırma alanından oluşmaktadır. İlk alan, yeni bir hibrit bulanık tahmin yöntemi aracılığıyla AVM ziyaretçi sayılarının tahmin edilmesine odaklanmaktadır. Bu yaklaşım, hava koşulları, finansal piyasa verileri ve internet trendleri gibi çevresel faktörlerin ziyaretçi desenleri üzerindeki etkisini analiz etmek için makine öğrenmesi algoritmalarını bulanık çok kriterli karar verme teknikleriyle entegre etmektedir. Perakende sektöründe lokasyon analizi ve müşteri talebi tahmini alanındaki literatür taraması, makine öğrenmesi yöntemlerinin yanında çok kriterli karar verme tekniklerinin birlikte kullanıldığı hibrit yaklaşımların sınırlı olduğunu göstermektedir. Özellikle AVM'lerin ziyaretçi sayılarının tahmin edilmesinde bu tür hibrit yaklaşımlar neredeyse hiç uygulanmamıştır.

Geliştirilen hibrit bulanık tahmin yöntemi, makine öğrenmesi algoritmaları ile bulanık çok kriterli karar verme tekniklerinin entegrasyonuna dayanmaktadır. Bu entegrasyon, tahmin modellerinde kullanılacak özelliklerin (features) bulanık çok kriterli karar verme teknikleri ile belirlenmesi ve bu özelliklerin makine öğrenmesi algoritmaları ile ağırlıklandırılması şeklinde gerçekleştirilmiştir. İstanbul'un Anadolu yakasındaki üç AVM'den elde edilen veriler modelin doğrulanması için kullanılmış, yüksek tahmin doğruluğu ve genellenebilirliği gösterilmiştir. Veri seti olarak, Haziran 2018 - Ocak 2019 dönemine ait günlük veriler kullanılmıştır. Bu veriler, büyük ölçekli bir teknoloji

ve telekomünikasyon hizmetleri sağlayıcısının baz istasyonlarından elde edilen sinyal verileri, hava durumu verileri, finansal piyasa verileri ve internet trend verileri gibi çeşitli veri kaynaklarını içermektedir.

Çalışma sonucunda, geliştirilen hibrit bulanık tahmin yönteminin AVM ziyaretçi sayılarını yüksek doğrulukla tahmin edebildiği gösterilmiştir. Bu model, AVM yöneticilerine hava durumu, finansal piyasalar ve internet trendleri gibi çevresel faktörlerin ziyaretçi sayıları üzerindeki etkisini anlama ve bu faktörlere dayalı tahminler yapma yeteneği sunmaktadır. Bu tahminler, personel planlaması, enerji tüketimi yönetimi, güvenlik hizmetleri, temizlik ve bakım hizmetleri gibi operasyonel alanlarda daha etkin kaynak tahsisi yapılmasına olanak sağlamaktadır. Önerilen metodoloji diğer perakende lokasyonlarına da uygulanabilir niteliktedir.

İkinci araştırma alanı, makine öğrenmesi ile Küresel Bulanık TOPSIS metodolojisini birleştiren hibrit bir yaklaşım geliştirerek internet üzerinden perakendede teslimat optimizasyonunu ele almaktadır. Online perakende sektörü, hem operasyonel maliyetleri hem de müşteri memnuniyetini etkileyen kritik bir bileşen olan son kilometre teslimat optimizasyonunda önemli zorluklarla karşı karşıyadır. Bu yenilikçi entegrasyon, makine öğrenmesi modellerinden elde edilen SHAP (Shapley Additive exPlanations) değerlerini Küresel Bulanık Kümelere dönüştürerek, belirsizlik altında daha etkili karar vermeyi mümkün kılmaktadır.

Çalışmada, online perakende teslimat verilerini kullanarak siparişlerin başarılı teslimat olasılığını tahmin eden bir Stokastik Gradyan Artırma modeli geliştirilmiştir. SHAP analizi sonucunda, teslimat başarısını etkileyen en önemli faktörlerin paket sayısı, kurye taşıma süresi, sipariş hazırlama süresi ve işlem tutarı olduğu belirlenmiştir. Bu faktörlerin önem dereceleri, küresel bulanık kümelere dönüştürülerek, aylık teslimat planlaması için iki farklı senaryo değerlendirilmiştir: verimlilik odaklı çoklu paket teslimat stratejisi ve müşteri memnuniyeti odaklı anlık teslimat stratejisi.

Küresel Bulanık TOPSIS yöntemi kullanılarak yapılan analiz sonucunda, belirli bir ay için çoklu paket teslimat stratejisinin daha uygun olduğu belirlenmiştir. Bu bulgu, online perakendecilere, teslimat stratejilerini çevresel faktörlere, müşteri tercihlerine ve operasyonel kısıtlamalara göre optimize etmeleri için değerli bir çerçeve sunmaktadır. Karar modeli, teslimat maliyetlerini azaltırken, müşteri memnuniyetini artırmayı ve çevresel sürdürülebilirliği sağlamayı hedeflemektedir.

Üçüncü araştırma alanı, işlem maliyeti ekonomisi perspektifiyle ürün geliştirme süreçlerindeki kaynak tahsisi kararlarını incelemektedir. İşlem maliyeti teorisi, Coase (1937) tarafından öncülük edilen ve daha sonra Williamson (1975, 1981) tarafından geliştirilen, organizasyonların ne zaman faaliyetleri dahili olarak gerçekleştirmesi ve ne zaman harici olarak sözleşme yapması gerektiğini anlamak için bir çerçeve sağlar. Bu teori üç temel boyut belirler: belirsizlik, varlık özgüllüğü ve sipariş sıklığı. İşlem maliyeti ekonomisi, dış kaynak kullanım maliyetleri ile iç üretim maliyetleri arasındaki fark olarak tanımlanır.

Bulanık Analitik Hiyerarşi Süreci'ni işlem maliyeti teorisine uygulayarak, çalışma iç ve dış kaynak tahsisi kararlarını etkileyen maliyet faktörlerini belirlemekte ve önceliklendirmektedir. Bu araştırma, teknoloji şirketlerinin ürün veya proje geliştirme sırasında optimal kaynak tahsisi stratejisini (iç veya dış) belirlemelerine yardımcı olacak bir karar verme yöntemi geliştirerek literatürdeki önemli bir boşluğu ele almaktadır. Yenilikçi katkı, çeşitli maliyet faktörlerinin işlem maliyetleri üzerindeki etkisini sayısallaştıran, şirketlerin kaynak tahsisi hakkında daha bilinçli stratejik kararlar almalarına olanak tanıyan çok kriterli bulanık bir karar verme çerçevesinin oluşturulmasıdır.

Çalışma, dijital ürün lansmanlarında deneyime sahip büyük ölçekli teknoloji şirketlerindeki ürün yöneticilerinden uzman girişi içeren kapsamlı bir metodoloji kullanmıştır. Bu süreç boyunca, hem iç üretim hem de dış kaynak kullanım senaryolarında ortaya çıkan tüm maliyetler belirlendi, kategorize edildi ve Bulanık Analitik Hiyerarşi Süreci (FAHP) kullanılarak önceliklendirildi. Bulgular, altyapı ve geliştirme maliyetleri (%44) ile belirsizlik ve zaman maliyetlerinin (%31) bu kararları etkileyen en önemli faktörler olduğunu ortaya koymaktadır. Bu iki faktörü, süreç ve onay maliyetleri (%12), insan kaynakları maliyetleri (%9) ve destek faaliyetleri maliyetleri (%4) takip etmektedir.

Bu bulgular, teknoloji şirketlerine ürün geliştirme süreçlerinde iç kaynak veya dış kaynak kullanımı kararını vermek için sistematik bir çerçeve sunmaktadır. Şirketler, belirli bir projede hangi maliyet faktörlerinin daha önemli olduğunu değerlendirerek, daha bilinçli ve maliyet-etkin kararlar alabilirler. Örneğin, belirsizlik ve zaman maliyetlerinin yüksek olduğu projelerde iç kaynak kullanımı daha uygun olabilirken, altyapı ve geliştirme maliyetlerinin yüksek olduğu projeler için dış kaynak kullanımı daha avantajlı olabilir.

Bu üç çalışmanın ortak özelliği, veri analitiği ile bulanık mantık yaklaşımlarını entegre eden hibrit metodolojiler geliştirmeleri ve bu metodolojileri gerçek sektör verilerine uygulamalarıdır. Bu entegrasyon, hem tahmin doğruluğunu hem de karar verme süreçlerinin etkinliğini artırmakta, perakende sektöründe maliyet avantajı sağlama potansiyeli sunmaktadır.

Tez, makine öğrenmesi ile bulanık karar verme tekniklerinin metodolojik entegrasyonunu ilerletme, işlem maliyeti teorisini ürün geliştirme karar verme sürecine genişletme ve perakende karar süreçlerinde belirsizlik modellemesine yeni bir yaklaşım getirme yoluyla hem teoriye hem de uygulamaya katkıda bulunmaktadır. Önerilen çerçeveler, perakendecilere giderek karmaşıklaşan pazar ortamında operasyonları optimize etmek, maliyetleri azaltmak ve rekabet gücünü artırmak için pratik araçlar sunmaktadır. Bu çalışmanın sonuçları, perakende sektöründe veri analitiğinin potansiyelini tam olarak realize etmek ve dijital dönüşüm sürecinde rekabet avantajı elde etmek için değerli içgörüler ve araçlar sağlamaktadır.

1. INTRODUCTION

The retail sector, along with its sub-sectors, is one of the most dynamic and rapidly transforming areas of the economy (Rooderkerk et al, 2022; Mou et al, 2018,). This sector has been structurally divided into three main sub-sector groups: physical/offline retail, online retail, and hybrid retail models, where both can be used simultaneously (Gauri et al., 2021). With the acceleration of digital transformation, the amount and variety of data produced in the retail sector has increased significantly. Analyzing this data and its use in strategic decisions has become one of the most important elements of gaining a competitive advantage (Shankar, 2019). Studies in the retail sector span different areas, such as customer management, data analytics, inventory management, supply chain optimization, and cost management (Mou et al, 2018,).

Costs in the retail sector have been steadily increasing, creating significant pressure on profit margins and necessitating the development of cost-focused models. Despite various attempts to implement cost reduction strategies, achieving substantial and sustainable cost advantages remains challenging in the retail industry (Shankar, 2019). Many cost-focused initiatives fail to deliver their expected benefits due to improper implementation, insufficient data analysis, or inadequate understanding of the complex interrelationships between different cost factors. This thesis explicitly addresses this gap by focusing on data-driven approaches to cost advantage modeling in the retail sector.

Establishing and effectively operating data-based models correctly is strategically important in obtaining a competitive advantage for businesses in the retail sector. Today, some of the most important problems encountered while creating data-based models are the inability to select the right problem for the model, determining and collecting appropriate data for the selected problem, inability to analyze the data with correct methodologies, and finally, the inability to transform the established model into a scenario that will correctly productize it (Adu et al, 2022). These challenges lead to the potential benefits of data-based models not being fully realized and investments not providing the expected returns.

1.1 Transformation of the Retail Sector and the Growing Importance of Data Analytics

The retail sector is transforming significantly due to technological developments, changing consumer expectations, and global competitive pressure. Traditional retailers are starting to sell through digital channels while e-commerce businesses are turning towards physical stores. This transformation increases the importance of data analytics at every stage of retail operations (Seetharaman et al, 2016).

Data analytics is used in many areas such as deep understanding consumer behavior, demand forecasting, inventory optimization and management, pricing, supply chain management, and warehouse operations in the retail sector. With the development of big data technologies, the growth of machine learning algorithms, and the increase in data storage capacities to perform analyses has been possible for retail sector at a scale and depth that was not previously possible (Sagar, 2024).

In the retail sector, main purpose of the data analytics applications are increasing operational efficency, improving customer experience and optimizing costs. Above all the data analytics aim to provide following solutions: optimizing inventory, logistics costs and marketing expenses, improving personnel productivity, reducing energy consumption (Raji et al., 2024).

On the other hand, applying data analytics to the retail sector also brings various challenges through adapting process. Retailers may prevent from making data driven actions because of data quality issues, integration difficulties, lack of analytical capabilities, incompatible organizational culture, and ethical/legal constraints (Sagar, 2024). To overcome these challenges and evaluate the potential of data analytics, actors in retail sector should have a systematic approach.

1.2 Problems Encountered in Data-Based Modeling and the Need for Solutions

The development and implementation of data-driven models in the retail sector present various problems (Bradlow et al, 2017). Defining these problems correctly and developing appropriate solutions are critically important to realizing the potential of data analytics to deliver cost benefits (Seetharaman et al, 2016).

Selecting the right problem is the first problem when applying data-based modeling (Ozdemir et al., 2021). There are many potential application areas in the retail sector, but determining which problems will create the highest value can be challenging. Considering businesses' strategic priorities, operational challenges and market dynamics is systematic approach to problem identification needed.

Secondly, identifying and collecting appropriate data for the chosen problem is challenging. Retailers produce a wide range of data, but the quality, consistency and accessibility of data may not be sufficient. Therefore, additional efforts and structured methodology may be needed to choose correct data (Gedik, 2024).

Third, choosing appropriate methodologies and algorithms is critical for data analysis. Varied analytical approaches (machine learning, statistical analysis, optimization and simulation) may be appropriate for different problems (Martinez et al, 2021). Each approach brings its own advantages, limitations, and application requirements. Selecting and applying the appropriate analytical methodology directly affects the usability and performance of the model (Nagpal et al, 2023).

Fourth, adapting developed models into business processes is a integrating developed models into practical applications and business processes is a significant challenge. Due to organizational, cultural or technical barriers often may technically successful models be ineffective. While integrating models into business processes, user adaptation and model maintenance should be addressed systematically (Nagpal et al, 2023).

Lastly, identifying effective strategies for the productization and sustainable operation of developed models is a critical concern. Strategic choice, such as whether model development and deployment should be handled internally or outsourced, which technological infrastructures should be employed, and how scalability will be managed, have significant implications for long-term success. Economic theories, transaction cost theory, show a valuable lens through which these decisions can be optimized (Argyres & Nickerson, 2019).

To overcome these problems thoroughly the entire process in the retail sector a holistic methodological framework is needed.

1.3 Scope and Objectives of the Thesis

This thesis is designed to develop data-based cost advantage models in the retail sector and analyze the applicability of these models. Within the scope of the study, three main research areas are focused on:

- Data-based models that will provide cost advantages in physical retail
- Data-based models that will provide cost advantages in online retail
- Optimization of resource allocation decisions with technology companies in the productization process of data-based models

The research conducted within the scope of the thesis examines in detail the processes of identifying problems that will create cost advantages in the retail sector, preparing data sets that can solve these problems, establishing data models, and productizing these models. In this way, end-to-end data modeling that will create cost advantages with a holistic perspective on the retail sector is provided.

In the study's first phase, shopping malls, Turkey's largest physical retail marketplaces, were selected for the physical retail application. A new hybrid fuzzy prediction method based on the integration of machine learning and fuzzy multi-criteria decision-making techniques was developed to predict the number of visitors to shopping malls. This method enables the analysis of factors affecting shopping mall visitor numbers and makes highly accurate predictions based on these factors. Accurate visitor predictions help shopping mall managers optimize personnel planning, energy management, marketing strategies and operational planning, thus providing cost advantages.

In the study's second phase, the online grocery category, which delivers instant orders to homes, was addressed for the online retail application. A hybrid approach based on the integration of machine learning and fuzzy TOPSIS method was developed to optimize delivery processes in this category. This approach enables the analysis of factors affecting delivery success and the evaluation of different delivery strategies. Optimization of delivery processes offers significant cost advantages in terms of reducing logistics costs, more efficient use of vehicle and personnel resources, and increasing customer satisfaction.

In the third phase of the study, a framework based on multi-criteria fuzzy decision-making techniques with a transaction cost theory perspective was developed to optimize resource allocation decisions through technology companies in the productization process of the developed data-based models. This framework helps optimize decisions on whether data-based models should be developed and operated with internal or external resources. The right resource allocation strategy aims to minimize costs in the development and implementation process of models while maximizing the performance and sustainability of the models.

These three research areas present a holistic approach to developing and implementing data-based cost advantage models in the retail sector. The study covers both physical and online retail areas and addresses the entire process from model development to productization. This holistic approach is critical to fully realizing the potential of data analytics in the retail sector.

1.4 Methodological Approach and Innovative Aspects

This thesis adopts an innovative methodological approach to the development and implementation of data-based cost advantage models in the retail sector. Hybrid integration of machine learning and fuzzy logic techniques are the most important feature of this approach. With the help of this hybrid model, the prediction accuracy of the models and ability to make decisions under uncertain environment increases.

Machine learning algorithms can be used providing powerful tools in order to identifying complex patterns and relationships in retail sector data and predicting future values. In operational optimization, prediction of customer behavior and demand forecasting in the retail sector is widely used in regression, classification, clustering, and time series analysis. Despite all this, there are limitations in decision making under uncertainty with machine learning models. Because the effectiveness of machine learning models depends on data quality and quantity.

On the other hand, fuzzy logic and multi-criteria decision-making techniques provide practical solutions under uncertainty and take into account many criteria in a balanced way. Fuzzy logic is a powerful tool for creating a model likewise human thought structure in uncertain, ambiguous situations or when the information are fragmented (Kahraman, 2008).

In this thesis, it is aimed to develop more effective and efficient solutions to solve complex and uncertain problems by combining the strengths of machine learning with fuzzy logic and multi-criteria decision-making techniques in the retail sector. This hybrid approach has been applied in three different research areas:

- A hybrid fuzzy prediction method was developed to estimate the number of shopping mall visitors. This method uses fuzzy multi-criteria decision-making techniques to select and weight the features in the estimation model. Afterwards, the number of visitors is estimated with help of the machine learning algorithms. This integration of two methods increases the performance of model and also strengthens the explainabilitiy and interpretability.
- A hybrid approach based on the integration of machine learning and a spherical fuzzy TOPSIS model was developed for delivery optimization in online retail. In this approach, factors affecting delivery success were analyzed and the importance degrees of these factors were determined using machine learning. Then, with the help of the fuzzy TOPSIS method, different delivery strategies based on these factors were evaluated. This integration makes it possible to benefit from data-driven insights and expert knowledge in delivery decisions.
- A framework of integrating transaction cost theory and fuzzy analytic hierarchy process (FAHP) was developed for resource allocation decisions in the productization process. In this framework, cost factors which are based on transaction cost theory were determined, and FAHP evaluated their relative importance. This integration enables the systematic consideration of economic and organizational factors in resource allocation decisions.

The most important innovative aspect of these hybrid approaches is that they offer more comprehensive and robust solutions by integrating methodologies that are traditionally used separately. By combining the data-driven insights of machine learning models with the ability of fuzzy logic and multi-criteria decision-making techniques to make decisions under uncertainty, more effective solutions are developed for complex problems in the retail sector.

Another innovative aspect of this thesis is the development of a new methodology for transforming SHAP (Shapley Additive exPlanations) values obtained from machine learning models into spherical fuzzy sets. This methodology establishes a bridge between the explainability of machine learning models and uncertainty management in multi-criteria decision-making processes, thus enabling more transparent, interpretable, and reliable decisions.

1.5 Research Methodology and Data Sources

This thesis adopts a mixed methodology utilizing both qualitative and quantitative research methods. This approach enables a more comprehensive examination of complex and multidimensional problems in the retail sector.

The research methodology consists of three basic phases: problem definition and literature review, data collection and analysis and model development and evaluation. Different methodological approaches and techniques have been used in each phase.

In the problem definition and literature review phase, a comprehensive literature review was conducted on data analytics applications in the retail sector, cost advantage strategies, and transaction cost theory. This review includes academic articles, books, conference papers, industry reports, and case studies. The literature review identified research gaps and clarified the focus areas of the thesis.

Qualitative and quantitative data were used in the data collection and analysis phase. Daily visitor data obtained from base stations of three shopping malls on the Anatolian side of Istanbul, weather data, financial market data, and internet trend data were used to predict shopping mall visitor numbers. Order and delivery data, customer information, courier performance data, and environmental factors were analyzed to optimize online grocery delivery. Cost factors whereas product development processes and parameters which are affects resource allocation decisions were examined for transaction cost analyses.

Various techniques were used such as regression analysis, stochastic gradient boosting, SHAP values analysis, fuzzy analytic hierarchy process (FAHP), and spherical fuzzy TOPSIS in data analysis. Regression analysis and stochastic gradient boosting techniques were used to develop prediction models, while SHAP values analysis was used to increase the explainability of model outputs. FAHP and spherical fuzzy TOPSIS were used to manage uncertainty in multi-criteria decision-making problems and determine optimal decisions.

Hybrid approaches were designed and tested in the model development and evaluation phase. The shopping mall visitor prediction model integrates regression analysis and

FAHP. The optimization of online retail delivery model includes stochastic gradient boosting, SHAP values analysis, and spherical fuzzy TOPSIS. The framework of the transaction cost decision integrates transaction cost theory and FAHP.

The accuracy of prediction, interpretability of model, practical applicability and cost advantage potential are the various criteria which were used to evaluate models. Adjusted R^2, AUROC, precision, recall, and F1 score are the metrics which are used to measure the performance of prediction models. The effectiveness of decision models are tested with expert evaluations, sensitivity analyses, and case studies.

This mixed methodological approach has enabled a more comprehensive examination of complex problems and therefore integrating data-driven insights, expert knowledge,obtaining more reliable and applicable results in retail sector.

1.6 Contributions and Importance of the Thesis

This thesis provides a comprehensive framework for the development of data-based and cost-effective models in the retail industry. At the same time, this study makes effective contributions to industry practices and academic studies. The most important and notable contributions of this study are summarized as follows:

1.6.1 Theoretical contributions

- Machine Learning and Fuzzy Logic Integration: The thesis integrates machine learning with fuzzy logic and multi-criteria decision making techniques. This hybrid model increases the models' accuracy of prediction and the ability to make decisions under uncertainty, offering more effective solutions to complex problems in the retail sector.
- Methodology for Transforming SHAP Values into Spherical Fuzzy Sets: In this thesis, a methodology was developed that allows the SHAP values obtained from the machine learning model to be converted into spherical fuzzy sets. With this methodology, a model that can be more explainable, transparent and reliable decisions was developed.
- Application of Transaction Cost Theory to the Productization of Data-Based Models: This thesis develops a new perspective to integrate transaction cost theory into the productization process of data-based models. This framework

prioritizes transaction cost drivers using the fuzzy hierarchy method to optimize resource utilization.

- A Holistic Approach for Data Analytics Applications in the Retail Sector: This thesis provides a holistic framework from the development of data-based models to the productization of these models in the retail sector. This approach differentiates itself from more narrowly focused studies in the literature and provides a comprehensive roadmap for the full use of the potential of data analytics in the retail sector.

1.6.2 Practical contributions

- Hybrid Fuzzy Model for Shopping Mall Visitor Prediction: The study presents a hybrid fuzzy prediction model developed using real data to estimate the number of shopping mall visitors. This model helps industry players make more accurate operational and strategic decisions by allowing them to estimate the number of shopping mall visitors more accurately. In this way, cost optimization and revenue growth potential are evaluated more accurately.
- Machine Learning and Global Fuzzy TOPSIS Integration for Delivery Optimization: The study proposes an innovative approach to optimizing delivery processes in online retail based on transforming SHAP values into global fuzzy sets. This approach provides online retailers valuable insights and tools for optimizing delivery strategies, reducing logistics costs, and increasing customer satisfaction.
- Decision Framework for Resource Allocation Decisions in the Productization Process: The study provides a practical decision framework based on transaction cost theory and FAHP to optimize the internal or external resource use decision in the productization process of data-based models. This framework helps technology companies and retailers determine more systematic and data-driven resource allocation strategies.
- Sectoral Application Examples for Data-Based Decision Making: The study demonstrates practical applications of data-based models in the retail sector through real cases. These application examples guide retailers and technology

companies on how they can develop and implement similar models in their organizations.

These contributions are critical to fully realizing the potential of data analytics in the retail sector and providing cost advantages. The thesis fills methodological gaps in academic literature and offers practical solutions for sectoral applications. Additionally, integrating different disciplines, such as machine learning and fuzzy logic, encourages interdisciplinary research and opens new research areas.

1.6.3 Structure and content of the thesis

This thesis is based on the integration of three scientific papers, each with its own literature review and methodological framework. Taken together, these studies provide a holistic response to key methodological and practical gaps in the development of data-driven decision-support systems in the retail sector.

These papers consist of three main sections:

- Data application that will create cost advantages for physical retail
- Data application that will create cost advantages for online retail
- Analysis of which resource will realize the transformation of applications into products

In the second section, data-based models developed for physical retail and applications of these models in shopping malls will be addressed. This section explains the hybrid fuzzy prediction method developed to predict shopping mall visitor numbers in detail. The theoretical foundations of the methodology, application steps, and results of the experimental study conducted on three shopping malls in Istanbul are presented. Additionally, potential cost advantage of this model and practical applications are analyzed.

In the third section, data-based models that are developed for online retail and applications of the online grocery sector are examined. In this section, optimization in the delivery process with the hybrid approach based on the machine learning model and spherical fuzzy TOPSIS is examined. Particularly, the factors that affect delivery success are analyzed throughout the transformation of the two methodologies from SHAP values into spherical fuzzy sets. As a result of these, created decision model for different delivery strategies is explained. Additionally, the implementation of this hybrid method in the online grocery sector leads to cost optimization potential.

In the fourth section, the productization processes of the developed data-based models will be analyzed from the perspective of the transaction cost theory. This section presents the basic principles of transaction cost theory, its applications in productization processes, and the fuzzy analytic hierarchy process-based decision model developed for prioritizing the cost factors. In addition, the factors that affect the decision to use internal or external resources are detailed in this section. And also the framework of the developed model for the optimizing decisions are also mentioned and detailed.

This thesis aims to contribute to both academic literature and industry practices by creating a comprehensive framework for data-based cost advantage models.

2. DEVELOPMENT OF A HYBRID FUZZY MODEL FOR PREDICTING SHOPPING MALL VISITOR NUMBERS

Shopping malls are one of the most important components of the physical retail sector and accurately predicting shopping mall visitor numbers is critically important for operational planning, marketing strategies, and cost management. This section will present studies on the hybrid fuzzy model developed for predicting shopping mall visitor numbers.

A literature review in the field of location analysis and customer demand prediction in the retail sector shows that hybrid approaches using multi-criteria decision-making techniques alongside machine learning methods are limited. In particular, such hybrid approaches have been almost never applied to predict visitor numbers for shopping malls. To address this gap, this study improved a new hybrid fuzzy prediction method.

The hybrid fuzzy estimation method combined the two techniques machine learning algorithms with fuzzy multi-criteria decision making. This combination achieved by determination of the features which will be used in estimation models with fuzzy multi-criteria decision making techniques. And also this combination achieved weighting these features with help of the machine learning algorithms. For the hybrid model, three shopping malls on the Anatolian side of Istanbul were selected and the predicted number of visitors to these malls. As the dataset from June 2018 to January 2019, visitor rates are used. For creating the dataset, different data sources used such as signal data from base stations of telecommunications service provider, weather data, financial market data and internet.

This section is based on the paper "Ozdemir, C., Cevik Onar, S., Bagriyanik, S., Kahraman,C., Akalin, B., Oztaysi, B. 2021. Estimating Shopping Center Visitor Numbers Based On A New Hybrid Fuzzy Prediction Method, *Journal of Intelligent & Fuzzy Systems*, 42(7), 1-14".

The study shows the hybrid fuzzy prediction method can accuratly estimate the number of the visitors of shopping malls. This model is a beneficial tool for shopping mall executives and the sector players. With help of this tool sector players can take strategic actions based on the estimation of future number of visitors in a more certain environment. In addition , the proposed metholodogy also can apply to other retail locations. The study demonstrated that the developed hybrid fuzzy prediction method could accurately predict shopping mall visitor numbers.

2.1 Introduction

The number of shopping centers is one of the main key monetary indicators for a state. Competition increases in parallel with the enlarging of shopping mall amount; therefore, shopping malls begin to transform their products and marketing activities for making a difference in competition (Chebat et al, 2010). The enlarging shopping mall amounts lead to myriad shopping malls not to attain the required number of customers. This condition creates a crucial barrier for shopping malls to attain the required incomes (Şekerkaya & Cengiz, 2010). Since the beginning of the investment, the shopping malls' main objectives are establishing the shopping mall in the proper place, enticing customers to this place, and attaining required income (Hedhli et al, 2013; Cheng et al, 2005). The two most crucial barriers to attaining required income are not estimating the number of customers precisely and not enticing customers to the shopping mall (Zolfani et al, 2013). One of the biggest problems of shopping malls in attracting customers to their locations and earning maximum profit from these customers is that they cannot achieve the right customer management strategy (Zolfani et al, 2013). The most critical points of these strategies within the scope of the shopping center are to determine the target audiences for the selected shopping center, to find out the prominent segmentations for these target groups, to understand the needs of the main segmentations, and to communicate with these audiences through the right channels (Gauvreau, 2018; El-Adly, 2007). This situation appears that it is challenging to meet the right customer management strategy's needs with traditional marketing methods. For this reason, under augmenting data variety and volume thanks to digitalization, customer management strategies have become data-oriented.

In the latest years, with the enhancing significance of data in all business lines, the data's worth also began to enhance the shopping mall business line (Wang et al, 2018;

Fayyad et al, 1996). Shopping centers have started to utilize this data, which is produced and reserved in order to estimate the number of customers and required income for the coming days. Besides customer data, shopping malls have also started to collect environmental factors to consider their effects and include them in their analysis (Wesley et al, 2006). When the literature investigation was applied primarily for shopping malls, it was found that there were few publications on the selection of best locations for shopping malls with multi-criteria decision-making methods. There were few publications regarding the on-demand forecast of customer numbers, and almost there were no publications where the demand forecast of customer numbers was merged with location analysis. In this study, it is used by merging a multi-criteria decision-making technic to estimate the customer numbers by location analysis, which are crucial for the shopping mall business line and are not available in the literature. The location analysis was applied together with the large-scale technology and communications services provider company. The number of customers on a daily basis for the selected shopping malls was found via the signaling raw data at the locations (Eiselt & Marianov, 2015; Eiselt & Marianov, 2011). A multi-criteria decision-making technique was applied to assist the prediction of the customer numbers for shopping malls. This study has two key targets. The first target is genuineness, predicting the customer numbers for the selected location by merging multi-criteria decision-making techniques, location analysis, customer demand forecasting analysis, and machine learning methods that have not been appropriately studied in the literature. A new hybrid fuzzy prediction method has been developed to forecast customer behaviors. The second target is contributing to the strategic activities and marketing plans of the shopping malls to predict the customer numbers and analysis of customer attitude according to the environmental factors.

In the next part of the study, the literature review details and the case study's analytical details are discussed. The third section explains the steps of the new hybrid fuzzy prediction method. In Section Four, the real case analysis and the results of the real case analysis are given. In the last part, conclusion and discussion are included.

2.2 Background

Previous studies for shopping centers have generally been customer-based. Therefore, there is a need for a study that includes the estimation of the number of customers daily and examines the environmental factors affecting the number of customers. As of 2010, the total number of studies, including location analysis and customer demand forecast analysis for all literature sectors, is 169. If this analysis is narrowed further, the number of studies for all sectors decreases to 13 by combining location analysis, customer demand forecast analysis, and machine learning algorithm. When we review these 13 studies on a sectoral basis, the number of retail sectors decreases to 6, and the number of resources in shopping centers decreases to 1. Table 2.1 and 2.2 shows the literature review details. When we examine this research further, customers' faces are observed through camera images, and their behavior will be estimated (Huotari, 2015). Since the number of resources in this area is limited, the factors affecting the number of visitors to the shopping center are not fully known. In order to detect what affects the estimation of the number of customers for shopping malls, multi-criteria decision-making methods are applied to decide what feature is more important than the alternatives. MCDM helps the decision-maker to prioritize options among its alternatives. There is no study found in literature concerning the combination of MCDM and using machine learning methods for estimating the number of visitors to the shopping center to the best of our knowledge. This study aims to determine the features entering the estimation model over those resulting from the MCDM. In this study, machine learning algorithms, which can be used in location analysis and customer demand forecasting models, were tested. However, the data includes daily customer numbers and environmental information for the selected shopping mall, and location-based data sets are available instead of person-based data. In addition to these, all data types are numerical. Considering all these constraints, since the classification structure cannot be established, artificial neural networks, support vector machines, naïve Bayes, and decision-tree learning techniques cannot be carried out in this study. In light of this information and the literature review, it was found that the most appropriate machine learning method for the application was regression analysis.

Table 2.1 : Customer management strategy literature review

2010-2019			
Customer Strategy Models	**Total**	**Machine Learning (ML)**	**ML + Regression**
Location Analysis	17,400	1,380	509
Demand Analysis	19,000	1,620	878
Location + Demand Analysis	**169**	**13**	**8**

Table 2.2 : Customer management strategy sectoral literature review

2010-2019				
Customer Strategy Models with ML	**Retail**	**Banking**	**Telco**	**Mall**
Location Analysis	203	190	104	28
Demand Analysis	254	225	92	21
Location + Demand Analysis	6	1	1	1

2.3 Methodology

With the digital transformation, the number of data sources in the world has increased significantly. Parallel to this, data types generated by corresponding sources are also increasing rapidly. Along with the increase in data types, estimation studies for business problems have also increased. Machine learning models used for these prediction studies are trained with existing data types. However, since there are different types of data, it is an arduous task to select the data types that will train machine learning algorithms for the variables that are planned to be predicted. Choosing the correct data types that will boost the estimation work's performance is also very important; thus, for prediction models, the correct features should be selected by following a systematic procedure.

In this study, New Hybrid Fuzzy Prediction Method is introduced by combining the fuzzy analytic hierarchy process (FAHP) and machine learning model.

Features that will train the machine learning algorithm were determined by using the FAHP method. The first step of FAHP is to give an importance value to the features with respect to the Prioritization Scale table. After the prioritization process of features, the pairwise comparison matrix is established based on the a_{ij}. Moreover, the Fuzzy AHP Method was applied to find normalized weights of features.

Table 2.3 : Prioritization Scale

Value	Numbers (a_{ij})
Equal	(1,1,1)
Moderate	(2,3,4)
Strong	(4,5,6)
Very Strong	(6,7,8)
Extremely Strong	(8,9,9)
Intermediate Values	(1,2,3)
	(3,4,5)
	(5,6,7)
	(7,8,9)

Step 1. Calculating fuzzy synthetic extent with respect to the i-th object is defined as (Chang, 1992):

$$S_i = \sum_{j=1}^{n} a_{ij} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \right]^{-1} \tag{2.1}$$

$$\begin{aligned} a_1 \oplus a_2 &= (l_1, m_1, u_1) \oplus (l_2, m_2, u_2) \\ &= (l_1 + l_2, m_1 + m_2, u_1 + u_2) \end{aligned} \tag{2.2}$$

$$(l_1, m_1, u_1)^{-1} = \left(\frac{1}{u_1}, \frac{1}{m_1}, \frac{1}{l_1} \right) \tag{2.3}$$

$$a_1 \otimes a_2 = (l_1, m_1, u_1) \oplus (l_2, m_2, u_2)$$
$$= (l_1 * l_2, m_1 * m_2, u_1 * u_2) \tag{2.4}$$

Step 2. Calculating the degree of possibility (Chang, 1992):

$$M_1 = (l_1, m_1, u_1)$$

$$M_2 = (l_2, m_2, u_2)$$

$$V(M_2 \geq M_1) = sup_{y \geq x}\left[\min\left(\mu_{M_1(x)}, \mu_{M_2(y)}\right)\right] \tag{2.5}$$

$$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2(d)}\ (see\ \boldsymbol{Fig.2.})$$
$$= \begin{cases} 1, & if\ m_2 \geq m_1 \\ 0, & if\ l_1 \geq u_2 \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & otherwise \end{cases} \tag{2.6}$$

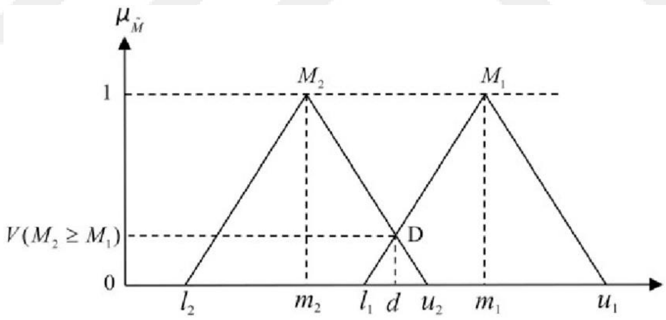

Figure 2.1 : The Intersection between M_1 and M_2.

Step 3. The degree possibility for a convex fuzzy number to be greater than k convex fuzzy numbers: (Kahraman et al, 2004)

$$V(S \geq S_1, S_2, \dots, S_k) = \min(S \geq S_i), \qquad i = 1,2, \dots, k \tag{2.7}$$

Step 4. Calculating weight vector:

$$d'(A_i) = min\, V(S_i \geq S_k), \quad i, k = 1,2, \ldots, n; k \neq i \tag{2.8}$$

$$W' = \left(d'(A_1),\ d'(A_2), \ldots,\ d'(A_n)\right)^T \tag{2.9}$$

Linear Regression algorithm was used to predict the number of customers of shopping malls and features, which was the output of the FAHP method used to train the algorithm. The following formulas explain the multiple regression analysis and cost function (2.10)-(2.11)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \in \tag{2.10}$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\mathrm{h}\left(x^{(i)}\right) - y^{(i)}\right)^2 \tag{2.11}$$

Adj R2, Success Rate, VIF, Durbin- Watson (Auto Correlation) were used to evaluate the efficiency of the prediction model, whose definition is described as the following formula (2.12)-(2.15).

$$R^2_{adj} = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1}\right] \tag{2.12}$$

$$R^2 = 1 - \frac{SS_{RES}}{SS_{Tot}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \tag{2.13}$$

$$Success\ Rate = \text{Avg.}\left(\frac{1 - |Prediction - Actual|}{\text{Actual}}\right) \tag{2.14}$$

$$VIF = \frac{1}{1-R_i^2} \qquad DW = \frac{\sum_{t=2}^{T}((e_t - e_{t-1})^2)}{\sum_{t=1}^{T} e_t^2} \tag{2.15}$$

There should be more metrics to evaluate the results of the machine learning algorithms to strengthen estimation models. Since MSE and F-test are very useful for comparing the models, they were used to understand the algorithms' performance. MSE and F-Test definitions are described as the following formula (2.16)-(2.17).

$$F = \frac{(SS_{Total} - SS_{Res})/(k-1)}{(SS_{Res})/(n-k)} \tag{2.16}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.17}$$

2.4 Case Study

In line with the application's main field, big data of large-scale technology and telecommunications services provider of telco company will be analyzed extensively. Since signaling data constitutes the majority of the big data in large-scale technology and telecommunications services provider, the data set that will enable the estimation study will be created over the signaling data of telco company. In this study, three shopping malls will be selected from the Anatolian side of Istanbul and estimation process of the number of the visitors will be applied to them. The time range for the data set which trains the machine learning model is determined as the period between June 2018 and January 2019. The number of individual customers will be determined based on the 217 day signaling data at the shopping mall location on a daily basis. According to data collected from indoor base stations, the average number of visitors per day of first, second and third shopping malls are 15.710, 6.455, 10.600 respectively. Utilizing from indoor base stations in shopping malls during data collection process increases the reliability of data used in models since indoor base stations are adjusted to serve only the places they are located in. Together with the obtained environmental data, the number of individual customers in the past period will be analyzed, and the number of customers going to the selected shopping mall will be predicted when environmental conditions change in the coming days. In this way, it will be possible to observe which variables affect the number of customers going to the shopping mall, which environmental variables are meaningful, and how the customer's changes.

2.4.1 Data Structure

Although the processing of unstructured and not clean data was challenging to get the data within this study's scope, the data was collected and made ready for analysis with data cleaning methods. Within this application's scope, the number of individual customers on a daily basis for selected shopping malls was achieved by leveraging the analysis ability of the large-scale technology and communications services provider company's signaling data. Table 2.4 shows the variable categories and data types in detail.

Table 2.4 : Variable list and details

Category of Variables	Variables	Data Type
Total Number of Customer	1	Output
Daily Temperature	3	Input
Daily Amount of Rainfall	1	Input
Financial Market Data	5	Input
Hourly Average Traffic Density	4	Input
Daily Average Traffic Density	1	Input
Internet Trend	4	Input

2.4.2 Data Evaluation with multi-criteria decision-making method

Data evaluation is one of the most challenging parts of the data-driven solutions for business problems. It is difficult to determine which feature has the priority to solve the problem since many data sources generate different features. Where feature importance cannot be evaluated simply, the multi-criteria decision-making method helps on prioritizing features. MCDM methods help decision-makers follow systematic procedures while prioritizing the adequate features under uncertain situations (Oguztimur, 2011).

There are mainly two groups, which are Multi-Objective Decision-making (MODM) methods and Multi-Attribute Decision-making (MADM) methods in the MCDM approach (see Fig. 2.2.). MADM methods have the advantage of being able to evaluate various options based on various criteria with different units and try to determine the best option among them, to rank or classify options from best to worst (Oguztimur, 2011).

In MADM methods, several types of fuzzy technics can be applied to several sectoral decision processes. Since the fuzzy technic brings a methodology to model ambiguity from the human perspective (Muñoz et al, 2017), the fuzzy MCDM method was used. Among the corresponding methods, the Fuzzy Analytic Hierarchy Process is flexible and powerful for making decisions. Therefore, within the scope of the estimating shopping center visitor numbers, FAHP was used to assist feature selection for the estimation model.

There are seven categories in the variable list: Total Number of Customer, Daily Temperature, Daily Amount of Rainfall, Financial Market Data, Hourly Average Traffic Density, Daily Average Traffic Density, and Internet Trend. FAHP was applied to the variable list to prioritize the features before putting the estimation model's data. Some Categories has less than three variables. Hence, the FAHP method was not applied to the corresponding categories. On the other hand, FAHP was implemented for categories having more than three variables.

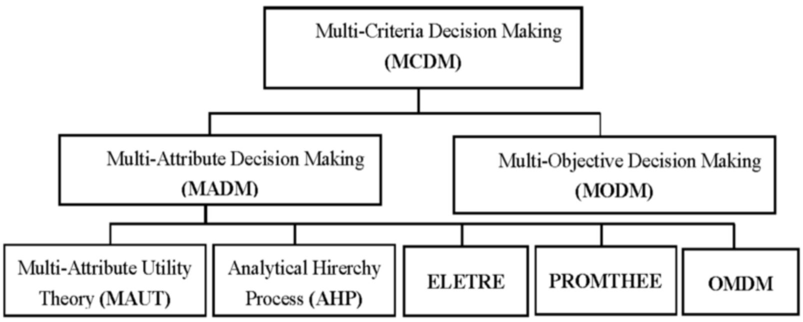

Figure 2.2 : Overview of the MCDM methods.

Steps mentioned above (2.1)-(2.9) were followed to analyze data used for the machine learning model. Prioritization scale was applied to Category of Variables, and subattributes were evaluated with the Fuzzy AHP Method.

Table 2.5 : Pairwise comparison matrix of category of variables

	DT	DAR	FMD	HATD	DATD	IT
DT	(1,1,1)	(3,4,5)	(2,3,4)	(4,5,6)	(1,2,3)	(1,1,1)
DAR	$(\frac{1}{5},\frac{1}{4},\frac{1}{3})$	(1,1,1)	$(\frac{1}{5},\frac{1}{4},\frac{1}{3})$	$(\frac{1}{3},\frac{1}{2},1)$	$(\frac{1}{5},\frac{1}{4},\frac{1}{3})$	$(\frac{1}{7},\frac{1}{6},\frac{1}{5})$
FMD	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	(3,4,5)	(1,1,1)	(3,4,5)	(1,1,1)	$(\frac{1}{3},\frac{1}{2},1)$
HATD	$(\frac{1}{6},\frac{1}{5},\frac{1}{4})$	(1,2,3)	$(\frac{1}{5},\frac{1}{4},\frac{1}{3})$	(1,1,1)	$(\frac{1}{5},\frac{1}{4},\frac{1}{3})$	$(\frac{1}{7},\frac{1}{6},\frac{1}{5})$
DATD	$(\frac{1}{3},\frac{1}{2},1)$	(3,4,5)	(1,1,1)	(3,4,5)	(1,1,1)	$(\frac{1}{3},\frac{1}{2},1)$
IT	(1,1,1)	(5,6,7)	(1,2,3)	(5,6,7)	(1,2,3)	(1,1,1)

In Table 2.5, the Pairwise Comparison Matrix was created for Category of Variables, and weights of criteria were calculated based on the corresponding matrix. The aim was to decide which category is the best choice to estimate visitors' numbers to shopping malls. DT, DAR, …, IT represent the following Variables respectively:

- DT: Daily Temperature
- DAR: Daily Amount of Rainfall
- FMD: Financial Market Data
- HATD: Hourly Average Traffic Density
- DATD: Daily Average Traffic Density
- IT: Internet Trend

Table 2.6 : Pairwise comparison matrix of daily temperature

	DDT	DNT	DAT
DDT	(1,1,1)	(1,1,1)	(1,2,3)
DNT	(1,1,1)	(1,1,1)	(1,2,3)
DAT	$(\frac{1}{3},\frac{1}{2},1)$	$(\frac{1}{3},\frac{1}{2},1)$	(1,1,1)

In Table 2.6, the Pairwise Comparison Matrix is shown. The matrix was created for Daily Temperature, and the weights of criteria were calculated based on the corresponding matrices. DDT, DNT, DAT represent the following Variables, respectively:

- DDT: Daily Daytime Temperature
- DNT: Daily Night Temperature
- DAT: Daily Average Temperature

Table 2.7 : Pairwise comparison matrix of financial market

	MCC	UTP	ETP	GTP	DBP
MCC	(1,1,1)	(3,4,5)	(3,4,5)	(1,2,3)	(2,3,4)
UTP	$(\frac{1}{5}, \frac{1}{4}, \frac{1}{3})$	(1,1,1)	(1,1,1)	$(\frac{1}{4}, \frac{1}{3}, \frac{1}{2})$	$(\frac{1}{3}, \frac{1}{2}, 1)$
ETP	$(\frac{1}{5}, \frac{1}{4}, \frac{1}{3})$	(1,1,1)	(1,1,1)	$(\frac{1}{4}, \frac{1}{3}, \frac{1}{2})$	$(\frac{1}{3}, \frac{1}{2}, 1)$
GTP	$(\frac{1}{3}, \frac{1}{2}, 1)$	(2,3,4)	(2,3,4)	(1,1,1)	(1,1,1)
DBP	$(\frac{1}{4}, \frac{1}{3}, \frac{1}{2})$	(1,2,3)	(1,2,3)	(1,1,1)	(1,1,1)

In Table 2.7, the Pairwise Comparison Matrix is shown. They were created for Financial Market, and weights of criteria were calculated based on the corresponding matrices. MCC, UTP, …, DBP represent the following variables, respectively:

- MCC: Monthly CPI Change
- UTP: Daily USD/TRY Parity
- ETP: Daily EUR/TRY Parity
- GTP: Daily GA/TRY Parity
- DBP: Daily Bist Parity

After Pairwise Comparison Matrices were generated for corresponding Categories, normalized weights were calculated with (2.1)-(2.4). The following weights and normalized weights for each category are in ascending order:

Table 2.8 : Weights of category of variables

Feature	Weights	Normalized Weights
IT	0.31	0.29
DT	0.30	0.28
DATD	0.19	0.17
FMD	0.17	0.16
HATD	0.06	0.05
DAR	0.05	0.04

Table 2.9 : Weights of daily temperature

Feature	Weights	Normalized Weights
DDT	0,41	0,39
DNT	0,41	0,39
DAT	0,24	0,23

Table 2.10 : Weights of financial market

Feature	Weights	Normalized Weights
MCC	0,45	0,41
GTP	0,25	0,23
DBP	0,19	0,18
UTP	0,10	0,09
ETP	0,10	0,09

Features were evaluated by looking at the output of the FAHP. When analyzing Table 2.8, 2.9, 2.10, there can be observed that some of the features come into prominence. In estimation models, High-weighted variables were used more frequently than the others.

2.4.3 Review of Regression Analysis Assumptions

In order to apply regression analysis, the error value for each variable to be used in the model should be zero, and the variances should be equal. Error-values should have a normal distribution, and each error value should be independent of its descriptor.

Besides, there should be no linear connection between independent variables and dependent variables, and the coefficients of independent variables should be linear. In addition, variables should conform to the metric scale. These assumptions re tested with the following analysis and the model was established with the variables suitable for these assumptions (Hair et al, 1995).

Suitability for normal distribution

Histogram graphics were examined in order to examine the conformity of the variables to normal distribution (George & Mallery, 2003; Barbara & Linda, 2013). It is observed that the variables selected according to these histogram graphics conform to the normal distribution (Field, 2009). Among the variables, DBP, GTP, UTP, ETP parities, and shopping mall internet trend (SMIT) was suitable for distribution (see Fig. 2.3.-Fig. 2.8.). However, normalization was carried out for the distribution of the amount of rainfall (DAR), daily night temperature (DNT), daily daytime temperature (DDT), daily average traffic density (DATD) and E-Commerce internet trend (ECIT) (see Fig. 2.9.-Fig. 2.13.).

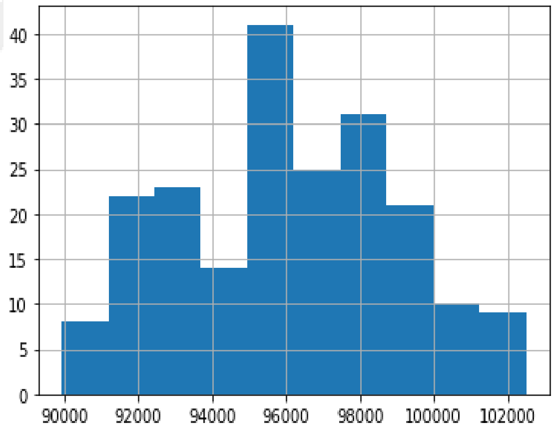

Figure 2.3 : Histogram of DBP

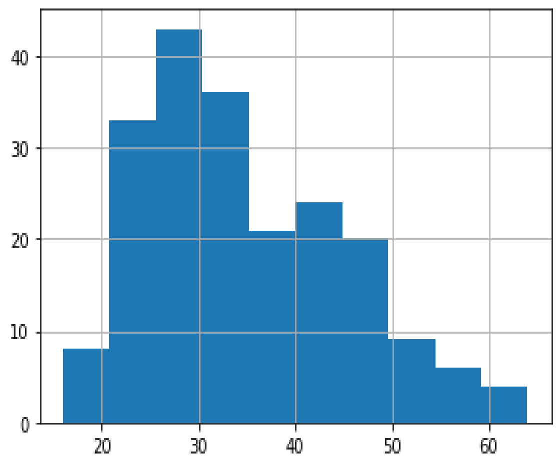

Figure 2.4 : Histogram of SMIT

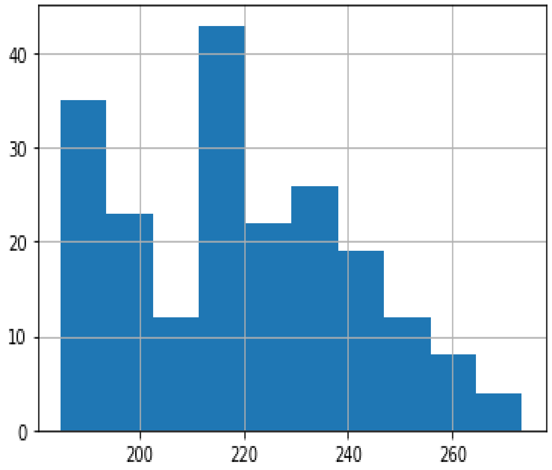

Figure 2.5 : Histogram of GTP

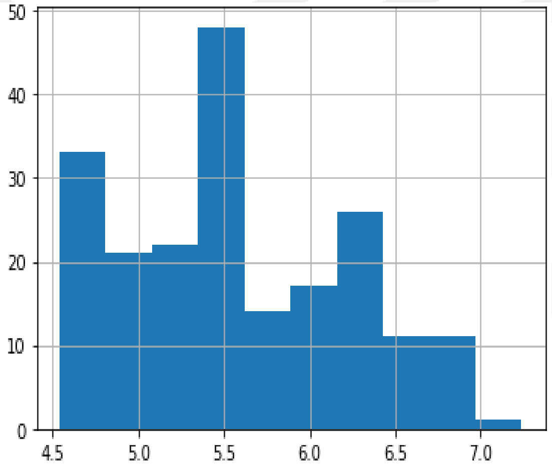

Figure 2.6 : Histogram of UTP

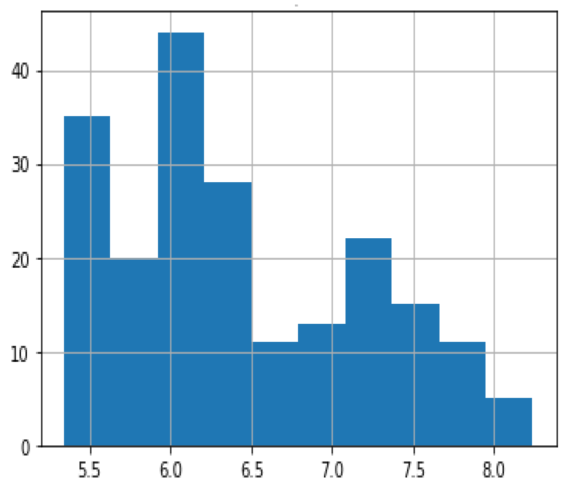

Figure 2.7 : Histogram of ETP

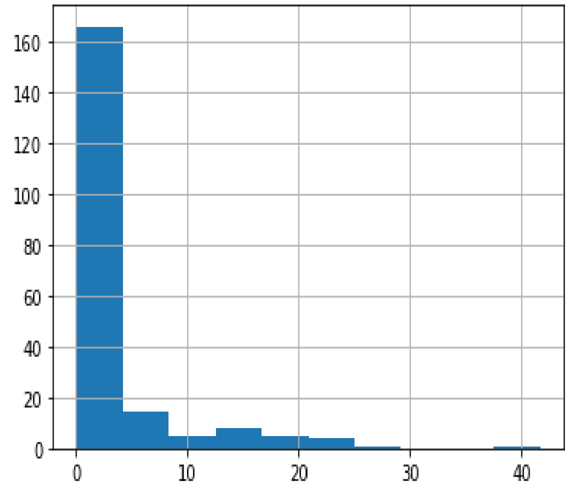

Figure 2.8 : Histogram of DAR

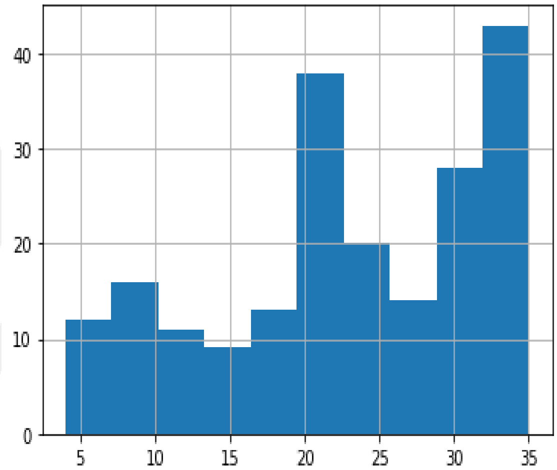

Figure 2.9 : Histogram of DDT

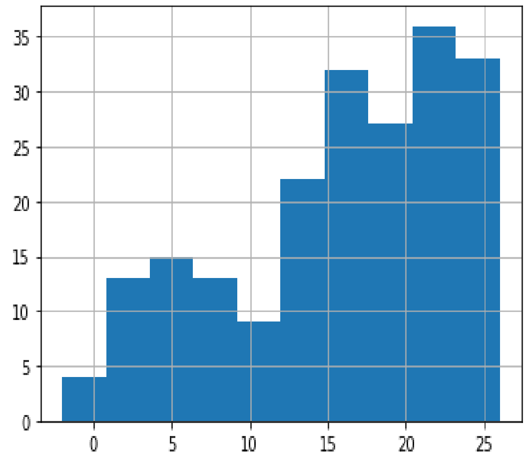

Figure 2.10 : Histogram of DNT

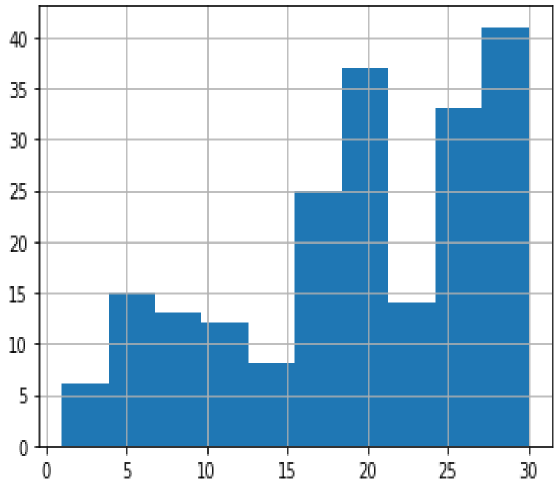

Figure 2.11 : Histogram of DAT

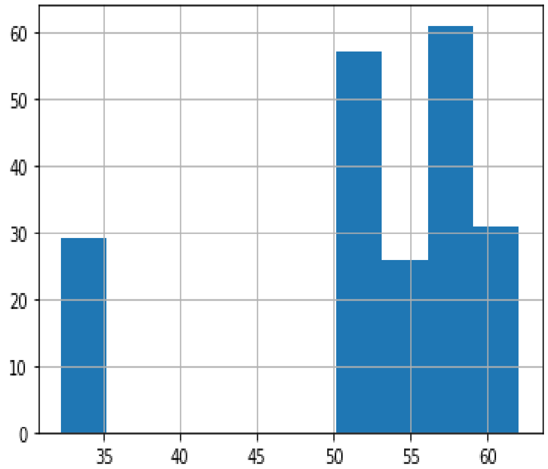

Figure 2.12 : Histogram of DATD

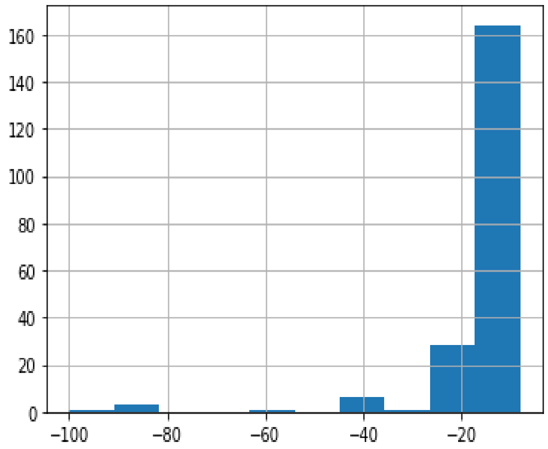

Figure 2.13 : Histogram of Daily ECIT

2.4.4 Data models for predictive analysis

The high-weighted features resulting from the above Fuzzy Analytic Hierarchy Process were used as inputs for the prediction models created. Since the importance of features with high weight is higher than the others, it was decided to use the corresponding features more frequently than low-weighted variables. In addition to the number of customers of the selected shopping center data in the past period, the total dataset was enhanced using the environmental and financial data. We want to predict the number of customers for the next day using machine learning algorithms on this data set. Linear regression analysis was the most appropriate model for our data set; therefore, Linear regression analysis was used as the machine learning algorithm. This study was predictive analysis because the aim was to predict the next day the number of customers. 20% of the data set has been used to test the model's performance, and 80% of the data set has been used for training the model for this machine learning algorithm. Multiple regression models were implemented to find out the best model that estimates the number of customers properly. Since it was aimed to find out which variable has the most notable effect on the estimation model, each regression model was prepared with different combinations of variables. In this context, by using random values, six regression analyzes were carried out, and the success rates were tested after the model setup.

2.4.4.1 First regression model

In the first regression model selected, SCIT, DNT, MCC, and DATD were selected as independent variables. The number of Customers (Total) was selected as the dependent variable. As a result of the machine learning and regression analysis, it was observed that the model was at a significant level with the F-test, and the R-adjusted was highly explainable with 95.7%. Besides, when Durbin-Watson and VIF values were examined, it was understood that these values were suitable for regression analysis.

As a result of this evaluation, four predictions were made for the first regression model, and it was determined that the prediction success rates were 98.5%, 98.7%, 95.6%, 99.8%, respectively, and the average success rate was 98.1%. This success rate and regression analysis F test also showed us that the selected independent variables SCIT, DNT, MCC, and DATD significantly affect the dependent variable (Number of Customers).

Table 2.11 : KPI comparison of regression models

Regression Models Name	Selected Variables	Adj R^2	Success Rate	Multi Collinearity	Auto Correlation
Model 1	SCIT, MCC, DNT and DATD	95.70%	98.10%	No	Positive
Model 2	SCIT, MCC, DNT, DATD and ECIT	95.70%	97.90%	No	Positive
Model 3	SCIT, GTP, DDT, DNT and DBP	95.40%	90.50%	Yes	Positive
Model 4	SCIT, DDT, DAR and DATD	95.60%	77.50%	No	Positive
Model 5	SCIT, DDT, DBP and MCC	95.60%	96.50%	Yes	Positive
Model 6	SCIT, DDT, DBP and GTP	95.40%	97.30%	Yes	Positive

2.4.4.2 Second regression model

In the second regression model selected, SCIT, DNT, MCC, DATD, and ECIT were selected as independent variables.The number of Customers (Total) was selected as the dependent variable.In this model, in addition to the first regression model, only ECIT was added to the model as an independent variable. As a result of the machine learning and regression analysis, it was observed that the model was at a significant level with the F-test and the model was highly explainable with 95.4% of R-adjusted. In addition, when Durbin-Watson and VIF values were examined, these values were suitable for regression analysis.As a result of this evaluation, four predictions were made for the second regression model, and it was determined that the prediction success rates were 97.9%, 98.2%, 96.2%, 99.3%, respectively, and the average success rate was 97.9%.This success rate and regression analysis F test also showed us that the selected independent variables SCIT, DNT, MCC, DATD, and ECIT had a significant effect on dependent variables (Number of Customers). However, when compared with the first model, when we look at the F-test value, VIF Value, Durbin-Watson, R-adjusted, and success rate parameters, the analysis result was slightly weakened.

This situation showed us that the first model was more effective than the second model.

2.4.4.3 Third regression model

As independent variables, SCIT, GTP, DDT, DNT, and DBP were selected in the third regression model. The number of Customers (Total) is selected as the dependent variable. VIF values were found to be very high due to the Collinearity test, and it was revealed that there was a linear relationship between the independent variables.

For this reason, the regression analysis was not continued, and only one estimation was made. According to this estimation result, the success rate was 90.5% and was lower than the first two regression models.

The model failed both in terms of success rate and exceeding statistical assumptions.

2.4.4.4 Fourth regression model

In the fourt regression model selected, SCIT, DDT, DAR, DATD were selected as independent variables.The number of Customers (Total) is selected as the dependent variable.As a result of the machine learning and regression analysis, it was observed that the model was at a significant level with the F test, and the R-adjusted was highly explainable with 95.4%. Besides, when Durbin-Watson and VIF values are examined, these values are suitable for regression analysis.As a result of this evaluation, the first estimation for the fourth regression model was produced, and the prediction success rate was 77.5%, and the estimation was stopped because this rate is far below the first three regression analysis models.

This success rate and regression analysis F-test also showed us that the selected independent variables SCIT, DDT, DAR, and DATD significantly affected dependent variables (Number of Customers).

However, when we compared it with the first two models, the success rate decreased significantly. This situation showed us that the first two models were more effective and successful than the fourth model.

2.4.4.5 Fifth regression model

In the fifth regression model selected, SCIT, DDT, DBP, MCC were selected as independent variables. The number of Customers (Total) was selected as the dependent variable.VIF values were found to be very high due to the Collinearity test, and it was revealed that there was a linear relationship between the independent variables.For this reason, the regression analysis was not continued, and only one estimation was made. According to this estimation result, the success rate was 96.5% and was lower than the first two regression models.The model failed both in terms of success rate and exceeding statistical assumptions.

2.4.4.6 Sixth regression model

In the selected sixth regression model, SCIT, DDT, DBP, and GTP were selected as independent variables. The number of Customers (Total) was selected as the dependent variable.VIF values were found to be very high due to the Collinearity test, and it was revealed that there was a linear relationship between the independent variables.For this reason, the regression analysis was not continued, and only one estimation was made. According to the result of this estimation, the success rate was 97.3%. Although the success rate is high, the model is not applicable due to the very high VIF rates of independent variables.

2.4.5 Data models evaluations

The six linear regression analysis created within the scope of the prediction analysis mentioned above shows us; The VIF rate is very high when at least two of the financial data and at least two of the temperature data are included in the same analysis. This reveals that the linearity between the variables is high. Considering the R-adjusted, F test, statistical estimation, and success rate, the first and second regression analysis turn out to be the most effective analyzes. When variables were normalized, it was observed that the error rate in the first and second regression analyzes decreased to a minimum and acceptable level. According to the above-mentioned scopes, it reveal (model 1) SCIT, DNT, MCC, DATD should be used in order to best estimate the customer numbers of ongoing days.

2.5 Strengthening data models

This study's primary purpose was to test the first and second regression analysis's scope and applicability based on the machine learning algorithm for shopping malls. In order to measure the validity of these tests, the models mentioned above were applied to both two other shopping centers. When the results of the regression analysis tests for new shopping malls were examined, it was revealed that the model was reasonable with the F-test and high R-adjusted results. Moreover, if we examine in more detail, these results also revealed the Suitability of Durbin-Watson and VIF values in the table below. Finally, comparing the models between different shopping malls, the success rate of 90% in both malls is observed. This rate represents a high success of the analysis. Although both regression models' success rate was higher for the first shopping mall, the error rate was higher for other shopping malls where the analysis was conducted. The fact that the success rate is higher in the first shopping center is that the model was established according to the first shopping center. Nevertheless, the model's application to the other two shopping centers and the acceptable error rate show us the success, validity, and dissemination of the regression analysis.When both regression models are examined, it is seen that the first regression analysis model is more successful in estimating. Table 2.12,2.13 shows the strengthening data model details below.

Table 2.12 : Strengthening details of first regression model

Model 1			
KPI	**First Shopping Mall**	**Second Shopping Mall**	**Third Shopping Mall**
Success Rate	98.10%	95.20%	93.70%
Adj R2	95.70%	96.60%	97.10%
MSE	0.82%	1.46%	1.63%
F Test	Significant	Significant	Significant
Multi Collinearity	No	No	No
Auto Correlation	Positive	Positive	Positive

Table 2.13 : Strengthening details of second regression model

Model 2			
KPI	**First Shopping Mall**	**Second Shopping Mall**	**Third Shopping Mall**
Success Rate	97.90%	94.90%	93.60%
Adj R2	95.70%	96.70%	97.10%
MSE	0.83%	1.45%	1.64%
F Test	Significant	Significant	Significant
Multi Collinearity	No	No	No
Auto Correlation	Positive	Positive	Positive

2.5.1 Testing models for the second shopping center

After the hypothetical tests, two outgoing data were found in the second shopping center internet trend variable, and the row with these two falling out data was deleted from the data set. A total of 217 daily data was transformed into 215 days of data with this deletion, and a new box diagram and exported data analysis were performed for the 215 days of data. The old and new box diagrams for the second shopping center internet trend are given below (see Fig. 2.14. and Fig.2.15.).

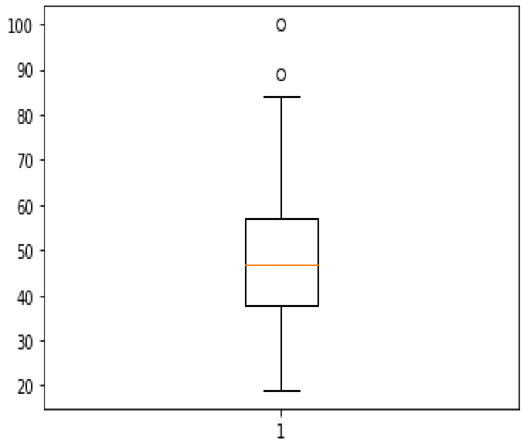

Figure 2.14 : Old box diagram of internet trend

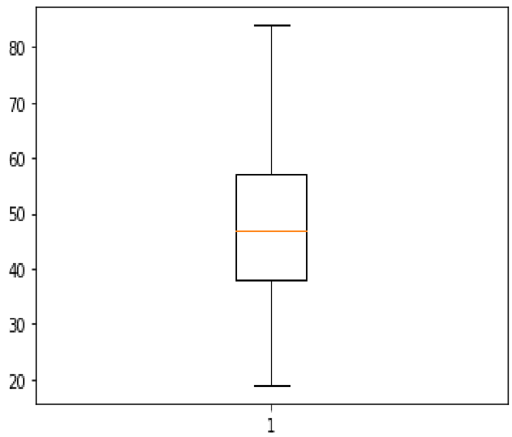

Figure 2.15 : New box diagram of internet trend

For the second shopping center's new data set, 172 days of data were used as train data (80%) and 43 days of data as test data (20%) in the first and second regression models.

It was observed that the machine learning and regression analysis performed in the first regression model selected, the model was at a significant level with the F test, and the R-adjusted was highly explainable with 96.6%. Besides, when the Durbin-Watson and VIF values are examined in the table below, these values were suitable for regression analysis. As a result of this evaluation, four predictions were made for the first regression model, and it was determined that the prediction success rates were 95.9%, 91.9%, 99.9%, 93.2%, respectively, and the average success rate was 95.2%. This success rate and regression analysis F-test also showed us that the selected independent variables SCIT, DNT, MCC, and DATD significantly affected dependent variables (Number of Customers).

The machine learning and regression analysis performed in the selected second regression model showed that the model was significant with the F-test, and the R-adjusted was highly explainable with 96.7%. When the Durbin-Watson and VIF values are examined, it was observed that these values were suitable for regression analysis.

As a result of this evaluation, four predictions were made for the first regression model, and it was determined that the prediction success rates were 95.1%, 92.2%, 99.5%, 93.0%, respectively, and the average success rate was 94.9%. This success rate and regression analysis F-test also showed us that the selected independent variables (SCIT, DNT, MCC, and DATD and ECIT) significantly affected dependent variables (Number of Customers).

Although the first and second regression models' average success rate in the second shopping center was reasonable, it was lower than the estimated average of the first shopping center selected. The established model was established according to the first shopping center's data and prepared as a suitable model. However, high estimation for the second shopping center, although not as much as the first one, revealed that these models could be applied to the second shopping center.

As in the first shopping center, the first regression model's prediction success rate was higher than the prediction success rate in the second regression.

2.5.2 Testing models for the third shopping center

After the hypothetical tests, 5 falling out data were found in the third shopping center internet trend variable, and the row with these five falling out data was deleted from the data set. With this deletion, a total of 217 daily data was transformed into 212 days of data, and a new box diagram and exported data analysis were performed for the 212 days of data (see Fig. 2.16. and Fig.2.17.).

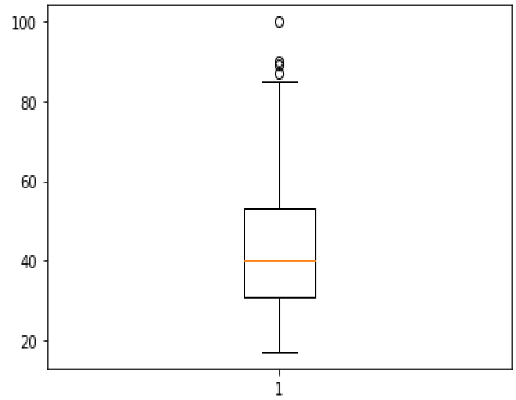

Figure 2.16 : Old box diagram of internet trend

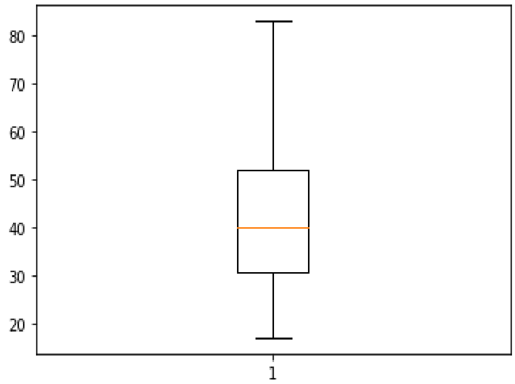

Figure 2.17 : New box diagram of internet trend

For the third shopping center's new data set, 169 days of data were used as train data (80%) and 43 days of data as test data (20%) in the first and second regression models.

It was observed that the machine learning and regression analysis performed in the first regression model selected, the model was at a significant level with the F test, and the R-adjusted was highly explainable with 97.1%. When the Durbin-Watson and VIF values were examined, it was stated that these values are suitable for regression analysis.

As a result of this evaluation, four predictions were made for the first regression model, and it was determined that the prediction success rates were 91.9%, 92.9%, 99.1%, 91.0%, respectively, and the average success rate was 93.7%. This success rate and regression analysis F-test also showed us that the selected independent variables SCIT, DNT, MCC, and DATD significantly affected dependent variables (Number of Customers).

The machine learning and regression analysis performed in the selected second regression model showed that the model was at a significant level with the F-test, and the R-adjusted was highly explainable with 97.1%. In addition, when the Durbin-Watson and VIF values are examined, it was stated that these values are suitable for regression analysis.

As a result of this evaluation, four predictions were made for the first regression model, and it was determined that the prediction success rates were 91.5%, 93.3%, 98.9%, 90.7%, respectively, and the average success rate was 93.6%. This success rate and regression analysis F test also showed us that the selected independent variables (SCIT, DNT, MCC, DATD, and ECIT) had a significant effect on dependent variables (Number of Customers).

If the result of forecasting the number of visitors for third shopping mall was considered alone, the average success rate of the first and second regression models were at an adequate level. However, if prediction result of third shopping mall was compared with prediction result of first shopping center, it was observed that third shopping center has lower success rate than the estimated average of the first shopping center selected. It was concluded that since the model was established according to the data set in the first shopping center, prediction result for the third mall was not as successful as the first one.

As in the first two shopping malls, the prediction success rate in the first regression model was higher than the prediction success rate in the second regression.

2.6 Discussion, Conclusion and Future Research

This study will benefit in the accurate estimation of the number of customers for the next days, one of the essential elements for shopping centers' success. The use of location analysis in estimating the customer numbers for shopping malls creates an important contribution to the literature. The fuzzy sets are excellent tools for representing the in human judgement (Kahraman et al, 2007; Kahraman et al, 2015; Oztaysi et al, 2018; Coban & Cevik Onar, 2018; Coban & Cevik Onar, 2017). The regressive models can be enhanced by using fuzzy sets. The evaluation of the data that will enter the estimation process with FAHP facilitated the feature selection. In addition to these, with this study, how the number of customer numbers of the shopping center changes according to the environmental climate, financial and internet trends, and the next day's estimation based on these variables are examined. According to the results of the analysis, it was found that the variables mentioned in the case study section had a significant effect on the number of customers. Furthermore, the model was found to be successful when applied to other shopping centers. This indicates the potential for the generalizability of the model.

This analysis was performed on a 217-day data set, and the results were based on these data. However, this analysis's validity may be reduced if the climate changes substantially, financial variables change in an unusual situation and if the internet trend is collected in a different format. For that reason, new variables, data sets, and new regression algorithms (SVM regression and ANN regression, etc.) may be needed to strengthen these estimation models in the subsequent analyzes.

In this study, the model and variables that should be used to estimate the number of customers for the next day with the help of the combined demand forecast model and location model of customer management strategies powered by a machine learning algorithm were examined. Using the FAHP method, the data that will enter the machine learning forecasting models enabled systematic procedure to be followed during the feature selection period. Regression analysis was found to be the most suitable machine learning technique for the model. The first and second regression models are the best models to estimate the number of customers within the analysis studies' scope. According to the machine learning models, the best environmental variables that estimate the number of customers for the next day are SCIT, MCC, DATD, and DNT. According to these results, by looking at environmental variables, shopping centers can predict the number of customers for the next day with a high success rate. To place these environmental variables in their models, they need to follow Istanbul's weather information, the daily/weekly / monthly change information of the finance market, traffic density information, e-commerce, and daily internet trend change. Since shopping centers will plan the next day with these models, they can provide both cost optimization and marketing strategies.

Moreover, shopping malls will be able to plan the number of security guards by predicting the number of incoming customers. When they reach a high number of customers, they will be able to increase the number of employees and create plans for the number of stores and consultants in the shopping center. Besides, shopping malls will be able to optimize their energy costs according to the estimated number of customers. In terms of marketing strategies, it will be possible to provide a good experience for new customers and to keep the existing customer loyal through campaigns. There will be an opportunity to change their strategic plans according to traffic density variability or market situation variability for shopping malls.

For future studies, the number of customers in all shopping malls should be examined. In order to deepen the scope of the analysis, the data should be enhanced. Different multi-criteria decision-making methods should be implemented on enriched data during the feature selection period. Simultaneously, other potential machine learning algorithms should be applied, and the success rate should be increased.

3. A HYBRID APPROACH TO DELIVERY OPTIMIZATION USING MACHINE LEARNING AND SPHERICAL FUZZY TOPSIS

The online retail sector faces significant challenges in the optimization of operational costs and customer satisfaction in the last mile delivery. This section shows innovative research that combines with machine learning prediction capabilities and advanced fuzzy decision-making techniques to optimize delivery processes. The research combines machine learning(ML), Spherical Fuzzy Sets(SFS) and the TOPSIS decision making methodology in addressing the complex problem of optimizing delivery success in online retail. This combination combines the predictive capability of the ML method with the uncertainty management capability of the fuzzy mind method, providing a more flexible and innovative decision-making approach.

Although ML models are quite good at predicting, they do not have the ability to manage uncertainties. The key innovation of this study is the transformation of SHAP (Shapley Additive Explanations) values into spherical fuzzy sets. With this transformation, the uncertainty in the decision-making process is detailed and a simpler decision-making process is created. The methodology of the research aims to collect and analyze data obtained from online retail delivery operations and then develop a Stochastic Gradient Boosting Model to predict delivery success. SHAP values are converted into fuzzy sets to determine the most important factors affecting delivery results. Finally, it is integrated with the TOPSIS method to address different distribution strategies. Two scenarios were examined: a multi-package delivery strategy focusing on efficiency and an instant delivery strategy prioritizing customer satisfaction. The scenarios were evaluated according to multiple criteria by taking into account the uncertainties in this scenario with the Spherical Fuzzy TOPSIS method. As a result, a monthly delivery and optimization framework was created.

This section is based on the paper “Ozdemir, C., Cevik Onar, S., Con, M. A Hybrid Approach to Delivery Optimisation Using Machine Learning and Spherical Fuzzy TOPSIS, Journal of Mult.-Valued Logic & Soft Computing (Accepted : December 2024).”

The research findings show us that with this hybrid approach, a more effective and reliable decision-making mechanism can be provided for optimization problems in complex delivery scenarios encountered in the online retail sector. Combining machine learning and fuzzy logic helps balance the conflicts of interest between customer satisfaction and operational efficiency. Balancing these problems ultimately results in providing cost advantage in the online retail sector.

This study represents a significant advance in data analytics applications for delivery optimization in the retail sector. It provides a methodological framework that industry players and researchers looking for solutions to improve logistics operations in the rapidly developing e-commerce sector can benefit from and develop.

3.1 Introduction

In recent years, the rapid growth of e-commerce and the proliferation of home delivery applications have led to significant changes in the logistics industry. This trend has led to the emergence of new needs and challenges, especially in last-mile delivery algorithms (Cote et al, 2018). One of the key factors determining online retail businesses' competitiveness is optimizing delivery routes and increasing delivery success (Jiang & Xinhui, 2020). The most important factor in ensuring these factors is the correct implementation and design of multivariate decision-making during order delivery. In business-to-consumer (B2C) retailing, the last-kilometer delivery of goods brings many challenges. Physical delivery is considered the most polluting and inefficient part of the logistics process (Gevaers et al, 2011). One of the most critical problems in this area is the existence of failed deliveries, incorrect decision-making, and correct decision-making mechanisms during decision-making. When unsuccessful order deliveries are retried, additional costs are incurred in package handling, fuel consumption and emissions. An important factor that influences decision-making is the presence of uncertainty. Studies have shown that successful deliveries increase customer satisfaction and reduce the probability of product returns (Rao et al, 2014).

Moreover, the environmental impacts of delivery operations are also becoming increasingly important. A study examined the impact of home delivery of grocery products on transport and CO2 emissions and considered the cooling factor (Heldt et al, 2019). Another research, based on case studies in Germany, the challenges and perspectives of using electric vehicles in e-commerce grocery deliveries were investigated (Ehrler et al, 2019). To address this complex and multifactorial problem, machine learning and artificial intelligence techniques have been widely used in recent years. One study developed a model-based approach for purchase forecasting in large product ranges (Jacobs et al, 2016). The stochastic gradient boosting method proposed by Friedman has significantly improved the performance of forecasting models (Friedman, 2002). Stochastic GBM is a method that can only be used to predict the success of an order. Although such models can make accurate forecasts, they may need to explain the factors driving them. To overcome this problem, we used the SHAP (Shapley Additive Explanations) method to explain which parameters influence the decisions of the GBM model and to what extent.

However, using more than these techniques is required to fully reflect the uncertainties and ambiguities in decision-making processes. Uncertainties encountered in the delivery process, unforeseen situations and errors in logistics planning can increase operational costs and negatively affect customer satisfaction. At this point, decision support systems such as machine learning and fuzzy logic play an essential role in the optimization of logistics processes. At this point, fuzzy logic, especially recently developed spherical fuzzy sets (SFS), stands out as a powerful tool for modeling uncertainty in decision-making processes (Kahraman & Kutlu Gundogdu, 2018).

The main objective of this study is to identify uncertainties more effectively with Spherical Fuzzy Sets (SFS) by using SHAP values obtained based on the previous machine learning model and to organize the multi-criteria decision-making process with the Spherical Fuzzy TOPSIS method. Spherical fuzzy sets are a model proposed by Kutlu Gundogdu and Kahraman that offers decision-makers a more comprehensive range of choices (Kahraman & Kutlu Gundogdu, 2018; Kutlu Gundogdu & Kahraman, 2019). This model allows us to define membership, neutrality and non-membership degrees independently. In spherical fuzzy sets, the sum of the squares of these three parameters can be at most 1, which provides decision-makers with more flexible modeling. This feature provides a more realistic representation of complex decision-making problems.

One of the most important advantages of spherical fuzzy sets is that the uncertainties encountered in classical fuzzy sets are handled in a wider scope. While the classical fuzzy logic proposed by Zadeh provides a powerful tool for managing uncertainties, extended models such as Intuitionistic Fuzzy Sets developed by Atanassov and

Pythagorean Fuzzy Sets proposed by Yager have introduced new methods on how to manage uncertainties in more complex situations (Zadeh, 1965; Atanassov, 1999; Yager, 2013). However, the scope of uncertainty in these models may remain limited. Spherical Fuzzy Sets consider both membership and non-membership degrees and the degree of uncertainty. Thus, they provide more accurate and reliable results in multi-criteria decision processes. The last of the approaches proposed in this study focuses on optimizing the multi-criteria decision-making process with the Spherical Fuzzy TOPSIS method based on the SHAP values obtained with the previous GBM model. TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method aims to reach the most appropriate solution by calculating the closeness of alternatives to the ideal solution in multi-criteria decision problems. Combining this method with fuzzy logic allows for more effective management of uncertainties and multiple criteria. In particular, the Spherical Fuzzy TOPSIS method allows uncertainties and hesitations to be handled more comprehensively, providing decision-makers with more reliable results. TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method is a multi-criteria decision-making technique developed by Hwang and Yoon (Hwang & Yoon, 1981). This method helps to determine the best option by evaluating the alternatives in terms of proximity to the ideal solution and distance from the negative ideal solution. Chen extended the TOPSIS method to the fuzzy environment, and then Boran et al. developed the intuitionistic fuzzy TOPSIS method (Chen, 2000; Boran et al, 2009). Kutlu Gundogdu and Kahraman developed the Spherical Fuzzy TOPSIS method by combining this method with spherical fuzzy sets (Kutlu Gundogdu & Kahraman, 2019).

This study will apply the Spherical Fuzzy TOPSIS method to two different scenarios for monthly delivery planning. The first scenario aims to evaluate a multi-package delivery strategy that aims to increase efficiency every month. In contrast, the second scenario will examine the monthly planning of an immediate delivery approach focusing on increasing customer satisfaction. These scenarios aim to identify the most appropriate monthly delivery strategies and optimize successful delivery rates under the influence of external factors such as order density, weather and traffic conditions. The Spherical Fuzzy TOPSIS method will help retail companies create optimal monthly delivery plans by balancing multiple criteria in this complex monthly decision-making process. In both scenarios, the parameters obtained with SHAP values will be modeled with fuzzy sets to manage the uncertainties encountered in the delivery processes more effectively and optimize the decision processes. This study presents a new approach to improve decision-making processes in last-kilometer delivery problems. In particular, by converting SHAP (Shapley Additive exPlanations) values obtained from machine learning models into spherical fuzzy sets, it is aimed to model the factors affecting delivery success in a more detailed and flexible way (Lundberg & Lee, 2017). This approach combines the predictive power of machine learning with the uncertainty modeling capability of spherical fuzzy sets.

The main contributions of this research can be summarised as follows:

- An innovative hybrid model is proposed, integrating machine learning and spherical fuzzy sets.
- A new methodology is developed to transform SHAP values into spherical fuzzy sets.
- A more comprehensive and flexible modeling of the factors affecting delivery success is provided.
- It provides an innovative model to minimize the uncertainty factor in delivery decisions. This model is blended with a machine learning model, SHAP value, spherical fuzzy set and topsis method.
- The membership degrees for spherical fuzzy sets and the decision matrices determined for TOPSIS are adapted from the parameter importance degrees affecting the successful delivery prediction model instead of expert opinion, which has the risk of being biased and is aimed to add innovation to this field.

This study aims to increase operational efficiency, improve customer satisfaction and develop sustainable logistics practices by providing more effective and efficient tools to decision-makers in logistics and supply chain management. Consequently, this study aims to extend the insights from SHAP values to optimize delivery processes in online retail by combining Spherical Fuzzy Sets and Spherical Fuzzy TOPSIS methods. This new approach aims to increase operational efficiency in logistics processes and maximize customer satisfaction by enabling more accurate management of uncertainties in multi-criteria decision-making problems.

The rest of the paper is organized as follows: Section 3.2 explains the background and literature review. Section 3.3 describes the research design in detail. Sections 3.4 and

3.5 present the materials and methods used. Sections 3.6 and 3.7 present the results obtained, while Section 3.8 discusses these results and suggests future research directions.

3.2 Background and Literature Review

In recent years, there has been an increasing need for improved methods to handle uncertainty and ambiguity in multi-criteria decision-making (MCDM) problems. In this context, various fuzzy set theories and decision-making methods based on them have become prominent, especially in complex and multidimensional problems. Fuzzy logic approaches have been effectively used in supply chain optimization problems to deal with uncertain factors such as costs and delivery times (Cevik Onar & Yasin Ates, 2007). Similarly, effective results were obtained by applying a fuzzy outranking method in facility location selection problems (Kaya & Cinar, 2007). In group decision-making problems, interactive decision support systems based on fuzzy logic have been developed to enable experts to express consistent preferences (Alonso et al, 2007). In supplier evaluation, fuzzy compromise ranking techniques were used to evaluate suppliers' environmental management performances (Buyukozkan & Feyzioglu, 2007). The importance of fuzzy decision-making and soft computing applications is increasing, and different fuzzy set theories are being used to develop new approaches to multi-criteria decision-making problems (Kahraman, 2008).

These developments show the applications of fuzzy set theories in various fields. However, classical fuzzy sets and intuitionistic fuzzy sets may be insufficient in some complex decision-making problems. In particular, when uncertainty is high, and decision-makers want to express their preferences more flexibly, existing theories are limited. To overcome these limitations and cover a broader spectrum of uncertainty, spherical fuzzy sets (SFS) and decision-making methods based on them have been developed (Kutlu Gundogdu & Kahraman, 2019; Kahraman et al, 2019; Kahraman & Kutlu Gundogdu, 2020).

In this context, spherical fuzzy sets (SFS) and decision-making methods based on them have become prominent, especially in complex and multidimensional problems. SFSs, as a generalization of classical fuzzy sets and intuitionistic fuzzy sets, offer a wider range of expressions to decision-makers (Kutlu Gundogdu & Kahraman, 2019).

Spherical fuzzy sets have membership (μ), neutrality (π) and non-membership (ν) degrees, and the sum of the squares of these three parameters can be at most 1: $0 \leq \mu^2 + \nu^2 + \pi^2 \leq 1$. This structure provides decision-makers with more flexible modeling, allowing more accurate results to be obtained, especially in situations with high uncertainty (Kutlu Gundogdu & Kahraman, 2019; Akram et al, 2021). These features of SFSs have made them an ideal tool for complex logistics and delivery problems such as last-mile delivery in online retail.

Spherical fuzzy sets and the spherical fuzzy TOPSIS method were introduced for the first time, as detailed by Kutlu Gundogdu and Kahraman, who also presented the definitions and operations of spherical fuzzy sets and the steps of the TOPSIS method (Kutlu Gundogdu & Kahraman, 2019). In addition, another study demonstrated the method's effectiveness by addressing the electric vehicle charging station location selection problem (Kutlu Gundogdu & Kahraman, 2020; Kahraman & Kutlu Gundogdu, 2020). This study provides a basic framework on how SFSs can be used in MCDM problems. This area has been researched in detail, and the literature survey has been extended. Another study in the literature emphasized the importance of aggregation operators in another study examining the applications of spherical fuzzy aggregation operators in multi-attribute group decision-making processes (Ashraf et al, 2019). Other studies on different application areas of SFSs suggest new approaches for decision-making and medical diagnosis problems. In this context, the potential applications of SFSs in the health sector have been examined, and the method's applicability in various fields has been demonstrated (Mahmood et al, 2019). At the same time, a study examining how SFSs are used in the field of sustainable transport evaluated vehicle technologies by combining spherical fuzzy AHP and TOPSIS methods (Jaller & Otay, 2021).

In addition, other studies detail the applications of SFSs in the transport and logistics sector (Thomas & Jose, 2023). These studies show how SFSs handle uncertain factors such as transport costs, supply and demand. In particular, the integration of fuzzy logic and TOPSIS in complex decision-making processes such as the transport of hazardous goods has been addressed as an important issue (Kanj et al, 2024). There are also studies showing the applicability of Pythagorean fuzzy set theory for transport management through a new distance measure (Sarkar & Biswas, 2021). Studies examining the applications of fuzzy set theory in supply chain management are also noteworthy. For example, studies using fuzzy logic theories in supplier selection emphasize the potential applications of SFSs in this area (Onat & Kactioglu, 2020). Other studies provide ideas on how SFSs can be used in hybrid (Duman et al, 2017). These studies examine how SFSs are used in sustainable supply chain management and how they can be applied to more complex MCDM problems (Alshammari et al, 2022). There have also been studies evaluating the application of SFSs in risk management (Ak & Demir, 2024). As a result, the literature has widely demonstrated that SFSs and the TOPSIS method can effectively reduce uncertainty in decision-making processes. These observations support the applicability of a hybrid spherical fuzzy and TOPSIS study in online retail delivery processes and route optimization.

In addition to these researches, there have been significant developments in areas such as the last-mile delivery success prediction model and route optimization in the online retail sector, which is the main topic of this article. The literature shows that route optimization studies have increased rapidly after 2001, while last-mile delivery research has intensified, especially after 2016. Although route, delivery and customer service problems are usually solved by classical optimization techniques such as traveling salesman algorithms, dynamically changing conditions require more flexible and powerful machine learning algorithms.

Research highlights the limited use of machine learning in route optimization, but there are more opportunities in this area in the future (Yuan et al, 2021; Li & Wang, 2020). For example, studies addressing how AI increases resilience in the supply chain and the importance of customer behavior analytics and optimal pricing strategies show how technology can improve operational performance (Singh et al, 2023). Similarly, studies addressing the e-commerce market from the circular economy perspective and big data analytics have revealed how technological advances can be integrated (Peng et al, 2022). Studies emphasizing the strategic advantages of big data analytics also provide essential findings (Ozemre & Kabadurmus, 2020).

Route optimization and last-mile delivery, where environmental factors and customer satisfaction are becoming increasingly important, are also gaining attention. Studies indicating that home delivery services contribute to reducing total vehicle kilometers and carbon emissions shed light on developments in this area (Heldt et al, 2019). Studies also examine the potential of automated delivery stations and strategies to reduce delivery errors (Oliveira et al, 2017) A study examining the use of home delivery services in Norway found that it encourages fewer trips to physical stores (Sund et al, 2021). Studies on expanding electric vehicle use in commercial transport provide essential insights into sustainable transport solutions (Ehrler et al, 2019). However, in all these studies, there are areas for improvement on the decision-making side, and if decision-making is added, the uncertainty dimension is omitted. To address this issue with an innovative approach, the hybrid model proposed in this paper offers a more effective and innovative solution by combining both the predictive power of machine learning methods and the flexibility of uncertainty management tools such as spherical fuzzy sets and TOPSIS. One of the most important innovations in the literature is the integration of SHAP values with spherical fuzzy sets. The use of SHAP values in this way has not been done before, and this study presents an innovative approach by converting SHAP values into spherical fuzzy sets, which better handles uncertainties and provides a more explanatory model in decision processes.

This model determines the spherical fuzzy set membership degrees and TOPSIS decision matrices using the importance degrees of the parameters affecting the delivery success prediction and eliminates the need for expert opinions. Thus, it minimizes the risk of expert bias and provides more objective and reliable results. Using SHAP values with such integration offers a significant innovation in reducing uncertainty and making decision-making processes more effective, especially in areas such as last-mile delivery, route optimization and delivery problems.

Furthermore, by enabling machine learning models to handle uncertainty better, this methodology aims to increase logistics efficiency and improve the accuracy of decision processes by providing flexible solutions in dynamic logistics processes such as delivery and route optimization. This hybrid approach provides a flexible decision-making solution path supported not only by machine learning but also by spherical fuzzy sets and TOPSIS methods.

3.3 Research Design

This study aims to predict the success of online retail deliveries and to establish a fuzzy-based decision-making structure based on the importance of each parameter resulting from this prediction as the primary objective. To achieve this, it uses a mixed-method novel approach that combines machine learning (ML) models with Spherical Fuzzy Sets (SFS). The research design begins with collecting and preparing a comprehensive dataset of variables related to online retail delivery outcomes. A rigorous cleaning and preprocessing process is applied to ensure data quality and consistency, followed by exploratory data analysis to understand the distribution and relationships of variables. Following data preparation, a stochastic gradient boosting model is developed and trained to predict delivery success. The model is validated using appropriate cross-validation techniques, and its performance is evaluated using relevant metrics such as accuracy, precision, recall and F1 score. SHAP (Shapley Additive ExPlanations) values are applied to identify the most effective factors in predicting delivery success. SHAP values are analyzed to understand the contribution of each attribute to the model's predictions.

The study's innovative aspect is the transformation of SHAP values into Spherical Fuzzy Sets for each criterion. This transformation paves the way for developing a Spherical Fuzzy TOPSIS model, which is used to evaluate different delivery scenarios and establish a multi-criteria decision-making structure. This integration provides a more comprehensive decision-making framework by combining the predictive power of ML models with the uncertainty modeling capability of SFSs. The research design involves developing a case study by applying the developed integrated ML-SFS approach to specific delivery scenarios. This application of the approach in real-world scenarios not only validates its effectiveness but also provides insights into its practical implications. In this case study, prediction accuracy and explainability results are analyzed. Furthermore, this framework will demonstrate the possibility of decision-making during delivery and afterward.

This research design aims to combine the predictive power of machine learning with the uncertainty-handling capabilities of Spherical Fuzzy Sets to create an innovative approach to improving decision-making in online retail delivery operations. This integrated approach effectively addresses complex operational challenges in the online retail sector by enabling business units to easily understand and mobilize complex predictive insights and explainability. As a result, this study aims to empower decision-making processes while providing a method that is robust in predictive capability and rich in explainability.

Methods:

- Data Collection and Preparation:
  - A comprehensive dataset of online retail delivery results was collected.
  - Cleaning and preprocessing were performed to ensure data quality and consistency.
  - Exploratory data analysis was performed to understand the distribution and relationships of variables.
- Machine Learning Model Development:
  - The Stochastic XGBoost model was developed and trained to predict delivery success.
  - The model was validated using appropriate cross-validation techniques.
  - The model's performance was evaluated using accuracy, precision, recall and F1 score metrics.
- Feature Importance Analysis:
  - SHAP values were applied to identify the most influential factors in predicting delivery success.
  - SHAP values were analyzed to understand the contribution of each attribute to the model's predictions.
- Spherical Fuzzy Set Integration:
  - SHAP values were transformed into Spherical Fuzzy Sets for each criterion.
  - Spherical Fuzzy TOPSIS model was developed to evaluate different delivery scenarios.
- Case Study Development:
  - The integrated ML-SFS approach was applied to specific delivery scenarios.
  - The results were analyzed for prediction accuracy and explainability

3.4 Materials and Methods – Machine Learning & Explainability

3.4.1 Data Sources and Preparation Process

In order to predict the delivery success of online retail orders, a comprehensive data set was initially collected, including a wide range of data on the ordering process. In line with the research objectives, this large dataset was adapted through a series of reduction processes, focusing on specific variables. Subsequently, a detailed exploratory data analysis (EDA) was performed on this organized dataset and cleaning processes were applied to ensure the dataset's quality (Ozdemir, Cevik Onar, & Bagriyanik, 2020). The organized dataset is enriched with data covering the entire order lifecycle, including detailed time information such as vendor waiting time, order assignment time, vendor acceptance time, order preparation and delivery times. In addition, environmental variables such as temperature and precipitation and binary indicators for days of the week are included in the dataset to consider environmental factors. These indicators allow us to analyze changes in order volumes and delivery efficiency. The most important variable, 'Target,' indicates the success of each delivery and allows classification models to predict this success. As a result, this dataset enables an in-depth understanding of the factors that influence the efficient completion of online retail orders.

3.4.2 Modelling and Data Processing Steps for Machine Learning

This study adopts a data-driven modeling approach to predict the successful delivery of online orders. Basically, the stochastic gradient boosting (SGB) algorithm is used to predict the delivery success of orders. However, several pre-processing steps were applied to the dataset to make the model more accurate, including missing data filling, data coding, feature scaling and methods to deal with class imbalance. Since missing data are rare, they were filled by simple averaging. The variables' values were scaled by the standardization method, also known as z-score normalization.

Data Encoding & Resampling

Two of the inputs (i.e., “day” and “daytime”) used in the predictive model are Categorical data such as 'day' and 'time of day' used in prediction models can only be used in machine learning analysis with quantification. For this reason, a one-hot coding method was applied so that each category was represented separately in the data. This process prevented the formation of artificial ordinal relationships. For example, the variable 'day' was sub-categorized from Sunday to Saturday, and the 'time of day' was reorganized into noon, afternoon, evening, and late evening.

Condensed Nearest Neighbour Under Sampling (CNN-US) method was preferred to eliminate data imbalance. This method eliminates the imbalance between classes and increases the sampling diversity in the data set (Ozdemir, Cevik Onar, & Ekmekcioğlu, 2024). This algorithm starts with a small sample set and then adds samples from the misclassified majority class. Although it is an algorithm sensitive to initial selections, CNN-US preserves the integrity of the data, ensuring that the basic information of the original data set is preserved.

Machine Learning Model

In this study, the SGB algorithm is preferred for performing classification tasks (Friedman, 2002). This algorithm, developed by Friedman, combines weak learners to obtain stronger models and aims to improve the model's accuracy through an iterative process. First, a decision tree model is traine d on the data set. Errors are calculated using the model outputs, and new decision trees are built on these errors. This process continues until the specified number of iterations is reached or the improvement in accuracy stops.

Performance Evaluation and Model Interpretability

For performance evaluation, the model's prediction accuracy was measured using AUROC. AUROC refers to the area under the ROC curve and evaluates the model's classification performance (Fawcett, 2006; Hand & Till, 2001). If the AUROC value is 0.5, it means that the model has no predictive ability, while the accuracy of the model increases as it approaches 1.

For the interpretability of the results of the model, the Shapley Additive exPlanations (SHAP) algorithm was used (Lundberg & Lee, 2017). The SHAP algorithm was preferred to understand better the reasons for the results predicted by the model and to increase the explainability of the results. This method provides in-depth insights into prediction analyses and clearly represents the effects of input variables on the target variable. It also increases the model's transparency by visually presenting each attribute's contribution to the model output.

3.5 Materials and Methods – Spherical Fuzzy Sets & Decision Making

Spherical Fuzzy TOPSIS method extend the TOPSIS by combining fuzzy logic. Spherical fuzzy environment can be used for representing a MCDM problem. The proposed spherical fuzzy TOPSIS method can be summarized in the following steps (Kahraman et al, 2019).

Step-1: Decision makers complete the decision matrices by using the linguistic terms given in Table 3.1. So, decision makers can define the weights by taking into account the corresponding spherical fuzzy numbers.

Table 3.1 : Linguistic terms and the corresponding spherical fuzzy numbers (Kutlu Gundogdu & Kahraman, 2019).

Linguistic Terms	Spherical Fuzzy Numbers (μ,ν,π)
Absolutely More Importance (AMI)	(0.9, 0.1, 0.1)
Very High Importance (VHI)	(0.8, 0.2, 0.2)
High Importance (HI)	(0.7, 0.3, 0.3)
Slightly More Importance (SMI)	(0.6, 0.4, 0.4)
Equally Importance (EI)	(0.5, 0.5, 0.5)
Slightly Low Importance (SLI)	(0.4, 0.6, 0.4)
Low Importance (LI)	(0.3, 0.7, 0.3)
Very Low Importance (VLI)	(0.2, 0.8, 0.2)
Absolutely Low Importance (ALI)	(0.1, 0.9, 0.1)

Step-2: Each decision maker's opinion combine by using Spherical Weighted Arithmetic Mean (SWAM) or Spherical Weighted Geometric Mean (SWGM) operators. Their formulas are seen below (Kutlu Gundogdu & Kahraman, 2019).

$$SWAM_W(\tilde{A}_{S1},\ldots,\tilde{A}_{Sn}) = \left\{\left[1-\prod_{i=1}^{n}(1-\mu_{\tilde{A}_{Si}}^2)^{w_i}\right]^{1/2}, \prod_{i=1}^{n}\nu_{\tilde{A}_{Si}}^{w_i}, \left[\prod_{i=1}^{n}(1-\mu_{\tilde{A}_{Si}}^2)^{w_i} - \prod_{i=1}^{n}(1-\mu_{\tilde{A}_{Si}}^2-\pi_{\tilde{A}_{Si}}^2)^{w_i}\right]^{1/2}\right\}$$

where $w=(w_1, w_2,\ldots, w_n)$, $w_i \in [0,1]$, $\sum_{i=1}^{n} w_i=1$. (3.1)

$$\mathrm{SWGM_W}(A_1,\ldots,A_n) = \left\{\prod_{i=1}^{n} \mu_{\mathrm{A_{Si}}}^{w_i}, \left[1-\prod_{i=1}^{n}(1-\nu_{\mathrm{A}_{Si}}^2)^{w_i}\right]^{1/2}, \left[\prod_{i=1}^{n}(1-\nu_{\mathrm{A}_{Si}}^2)^{w_i} - \prod_{i=1}^{n}(1-\nu_{\mathrm{A}_{Si}}^2-\pi_{A_{Si}}^2)^{w_i}\right]^{1/2}\right\}$$

where $w=(w_1, w_2,\ldots, w_n)$, $w_i \in [0,1]$, $\sum_{i=1}^{n} w_i=1$. (3.2)

Step-3: Each decision maker expresses their opinion on the criteria weights. Then, weights of the criteria are combined by using the SWAM operator.

Step-4: Aggregated weighted spherical fuzzy decision matrix is created by using the multiplication formula below.

$$\left\{\mu_{\tilde{A}_s}\mu_{\tilde{B}_s}, \left(\nu_{\tilde{A}_s}^2+\nu_{\tilde{B}_s}^2-\nu_{\tilde{A}_s}^2\nu_{\tilde{B}_s}^2\right), \left(\left(1-\nu_{\tilde{B}_s}^2\right)\pi_{\tilde{A}_s}^2+\left(1-\nu_{\tilde{A}_s}^2\right)\pi_{\tilde{B}_s}^2-\pi_{\tilde{A}_s}^2\pi_{\tilde{B}_s}^2\right)\right\} \quad (3.3)$$

Step-5: Aggregated weighted spherical fuzzy decision matrix is defuzzified by using the equation below and score function values are found.

$$\mathrm{Score}\,(\tilde{A}_s) = (\mu_{\tilde{A}_s} - \pi_{\tilde{A}s})^2 - (\nu_{\tilde{A}_s} - \pi_{\tilde{A}s})^2 \quad (3.4)$$

Step-6: Spherical Fuzzy Positive Ideal Solution (SFPIS) and the Spherical Fuzzy Negative Ideal Solution (SFNIS) is determined by taking into account the scores in the step 5.

Step-7: For each alternative, distances between alternatives and SFPIS and alternatives and SFNIS are calculated by using normalized Euclidean distance.

SFPIS:

$$D(X_i,X^*) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}\left(\left(\mu_{x_i}-\mu_{x^*}\right)^2+\left(\nu_{x_i}-\nu_{x^*}\right)^2+\left(\pi_{x_i}-\pi_{x^*}\right)^2\right)} \quad (3.5)$$

SFNIS:

$$D(X_i,X^-) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}\left(\left(\mu_{x_i}-\mu_{x^-}\right)^2+\left(\nu_{x_i}-\nu_{x^-}\right)^2+\left(\pi_{x_i}-\pi_{x^-}\right)^2\right)} \quad (3.6)$$

Step-8: Maximum distance to SFNIS and minimum distance to SFPIS are determined according to the result of the step 7.

Step-9: Closeness ratio is calculated by using the formula below.

$$\text{Closeness Ratio } (X_i) = \frac{D(X_i,X^*)}{Dmin(X_i,X^*)} - \frac{D(X_i,X^-)}{Dmin(X_i,X^-)} \tag{3.7}$$

Step-10: Rank the alternatives.

This method is designed to obtain more precise and reliable results in decision-making problems involving uncertainty and complexity. Spherical Fuzzy TOPSIS, while maintaining the advantages of the traditional TOPSIS method, allows decision makers to express their preferences more comprehensively thanks to the possibility of defining extended membership degrees offered by spherical fuzzy sets. With this feature, the method stands out as an effective tool especially in multi-criteria and complex decision-making problems.

3.6 Results – Machine Learning & Explainability

The main objective of this research is to predict the delivery success of orders. For this purpose, the stochastic gradient boosting (SGB) algorithm is integrated with the SHAP (Shapley Additive exPlanations) approach based on game theory. Within the scope of the SGB algorithm, the learning rate is set as 0.01, the minimum number of samples to be included in a leaf node is defined as 8, and the minimum number of samples to be included in a leaf node is defined as 5. In addition, the number of boosting steps to be applied was set as 250, and the maximum depth of each predictor was set as 3. The prediction model's performance was evaluated with the AUROC metric, and the related graphs are presented in Figure 3.1. The graph shows that the training performance of the model is relatively high, the AUROC value is 0.93 and the test results provide a satisfactory performance with an AUROC value of 0.88.

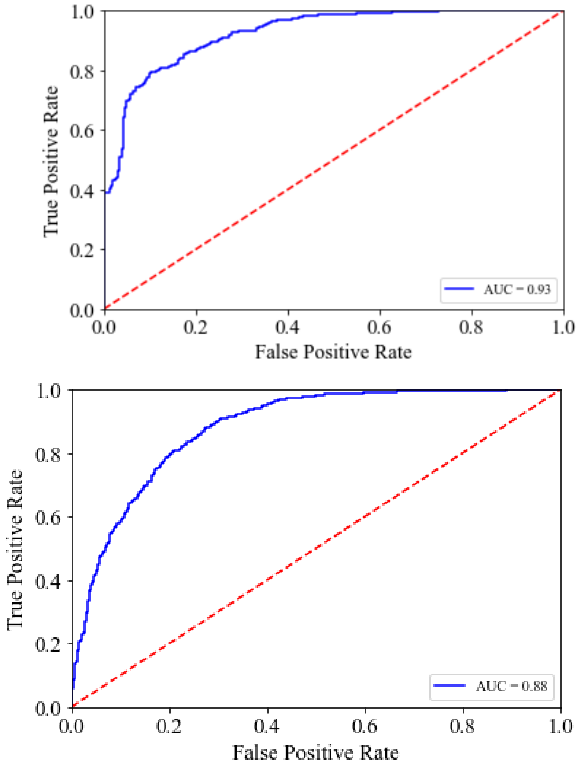

Figure 3.1 : Area under ROC plots a) Training b) Testing

This study also extracted the contribution of each feature to the predictions using the SHAP method, which is a model-independent algorithm. In this context, SHAP summary plots were created (Figure 3.2). In these graphs, Shapley values are shown on the horizontal axis, and feature names are shown on the vertical axis. In addition, the legend on the right side of the graphs represents different values with red and blue colors. Dark red dots represent high factor values, while dark blue dots symbolize low factor values. Positive Shapley values indicate delivery-compliant forecasts, while negative Shapley values indicate delivery failure. Furthermore, the higher a factor is on the graph, the stronger its influence on the model forecasts, as determined by the cumulative sum of the absolute Shapley values. A careful examination of the data shows that the factor with the largest impact on the predictions is the number of parcels, which indicates that it has the highest impact on the predictions. This finding indicates that the number of parcels significantly influences the probability of successful delivery of orders (Figure 3.2). The number of parcels is followed by courier transport time, order preparation time and processing time, and these factors significantly impact the delivery process. Higher parcel counts positively correlate with the probability of successful delivery, indicating a direct relationship between parcel volume and favorable outcomes. In contrast, long order preparation times negatively impact the delivery process's efficiency, risking delays or disruptions. In addition, the assignment time of the order also stands out as an important determinant, emphasizing the negative effects of long assignment times on delivery success. This highlights the critical importance of fast assignment planning to ensure the timely fulfillment of orders.

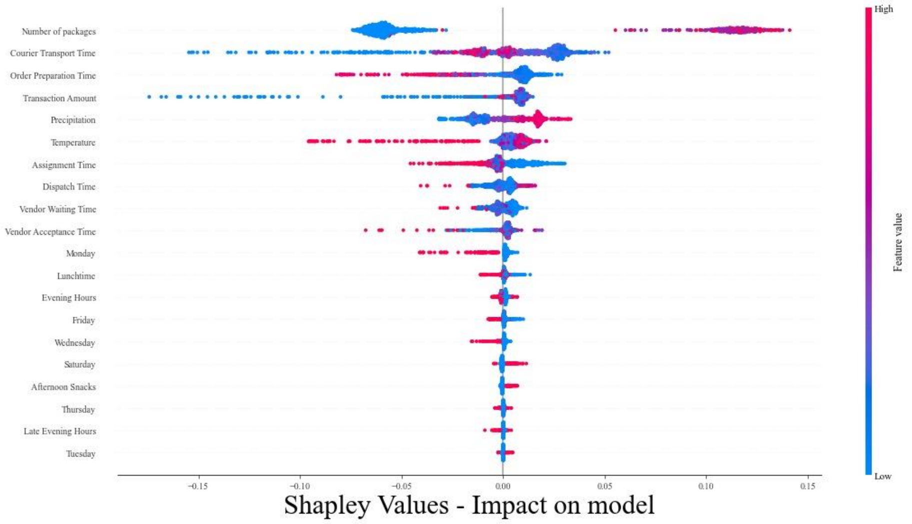

Figure 3.2 : SHAP summary plot

The following section will explain in detail how to connect and integrate these values to the spherical fuzzy set and TOPSIS.

3.7 Results – Spherical Fuzzy Sets & Decision Making

Today, the rapid growth in the e-commerce sector and increasing customer expectations require more effective and efficient decisions in logistics processes. In this context, it is critical for companies to evaluate the effectiveness of different delivery methods such as multi-package delivery strategy and immediate delivery approach. Aim of the study is to determine the optimal delivery strategy by using SHAP (SHapley Additive exPlanations) values and Spherical Fuzzy TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method with a hybrid approach.

In the research, an online retail dataset containing monthly supplier data was used. This data set includes 21 different criteria. The objective of the research is to decide whether this customer group should receive multi-package deliveries or immediate deliveries in the next month. The results of this decision will directly affect the arrangements to be made in operational areas such as warehouse organisation, staff training and resource management.

In the methodology, SHAP values obtained from the dataset are used to determine the criteria weights. These values are combined with the values in the decision matrix and Spherical Fuzzy TOPSIS method is applied. In this way, a decision problem with 21 criteria and 2 alternatives is analysed.

In the first step, a joint evaluation was obtained from three experts working in the operations unit in the online supply sector. This evaluation was made using predetermined linguistic terms given in Table 1 and the results are presented in Table 3.2. Under the assumption of a single decision maker, these evaluations were directly converted into Spherical Fuzzy numbers and the resulting decision matrix is shown in Table 3.3.

Table 3.2 : Evaluation of decision maker

Criteria	Multi-Package Delivery	Immediate Delivery
Number of packages	VHI	EI
Courier Transport Time	EI	HI
Order Preparation Time	LI	HI
Transaction Amount	HI	EI
Precipitation	EI	EI
Temperature	EI	EI
Assignment Time	EI	HI
Dispatch Time	LI	HI
Vendor Waiting Time	HI	LI
Vendor Acceptance Time	EI	HI
Monday	EI	EI
Lunchtime	EI	HI
Evening Hours	HI	HI
Friday	EI	HI
Wednesday	EI	EI
Saturday	HI	VHI
Afternoon Snacks	EI	HI
Late Evening Hours	LI	HI

Table 3.2 (continued) : Evaluation of decision makers

Tuesday	EI	EI
Sunday	HI	VHI

Table 3.3 : Decision matrix

Criteria	Multi-Package Delivery	Immediate Delivery
Number of packages	(0.8, 0.2, 0.2)	(0.5, 0.5, 0.5)
Courier Transport Time	(0.5, 0.5, 0.5)	(0.7, 0.3, 0.3)
Order Preparation Time	(0.3, 0.7, 0.3)	(0.7, 0.3, 0.3)
Transaction Amount	(0.7, 0.3, 0.3)	(0.5, 0.5, 0.5)
Precipitation	(0.5, 0.5, 0.5)	(0.5, 0.5, 0.5)
Temperature	(0.5, 0.5, 0.5)	(0.5, 0.5, 0.5)
Assignment Time	(0.5, 0.5, 0.5)	(0.7, 0.3, 0.3)
Dispatch Time	(0.3, 0.7, 0.3)	(0.7, 0.3, 0.3)
Vendor Waiting Time	(0.7, 0.3, 0.3)	(0.3, 0.7, 0.3)
Vendor Acceptance Time	(0.5, 0.5, 0.5)	(0.7, 0.3, 0.3)
Monday	(0.5, 0.5, 0.5)	(0.5, 0.5, 0.5)
Lunchtime	(0.5, 0.5, 0.5)	(0.7, 0.3, 0.3)
Evening Hours	(0.7, 0.3, 0.3)	(0.7, 0.3, 0.3)
Friday	(0.5, 0.5, 0.5)	(0.7, 0.3, 0.3)
Wednesday	(0.5, 0.5, 0.5)	(0.5, 0.5, 0.5)
Saturday	(0.7, 0.3, 0.3)	(0.8, 0.2, 0.2)
Afternoon Snacks	(0.5, 0.5, 0.5)	(0.7, 0.3, 0.3)
Thursday	(0.5, 0.5, 0.5)	(0.5, 0.5, 0.5)
Late Evening Hours	(0.3, 0.7, 0.3)	(0.7, 0.3, 0.3)
Tuesday	(0.5, 0.5, 0.5)	(0.5, 0.5, 0.5)
Sunday	(0.7, 0.3, 0.3)	(0.8, 0.2, 0.2)

When the values in the decision matrix are analysed, it is seen that Multi-Package Delivery has a higher value in the ‘Number of packages’ criterion. This is due to the fact that multi-package deliveries usually contain more packages. On the other hand, the Immediate Delivery option has higher values in the ‘Courier Transport Time’ and ‘Order Preparation Time’ criteria. This is due to the fact that immediate delivery is inherently faster. The ‘Transaction Amount’ and ‘Vendor Waiting Time’ criteria were considered more critical for Multi-Package Delivery as it is usually necessary to prepare large orders and were given higher values. For days of the week and time periods, scores were given according to the popularity of both delivery types. As an innovative approach, normalised SHAP values were used to determine the criteria weights. Since the SHAP value of the ‘Number of packages’ criterion was significantly higher than the others, this criterion was removed and the remaining criteria were normalised using the min-max normalisation method. ‘Number of packages’ criterion was assigned a normalization value as 1. SHAP values and normalized values are shown in Table 3.4. The normalised values were divided into 9 linguistic terms as shown in Table 3.5 and the Spherical Fuzzy number of each criterion was obtained. The results are presented in Table 3.6.

Table 3.4 : SHAP and normalised values of criteria

Criteria	SHAP Values	Normalised Values
Number of packages	0.0772	1.0000
Courier Transport Time	0.0244	1.0000
Order Preparation Time	0.0143	0.5817
Transaction Amount	0.0134	0.5430
Precipitation	0.0129	0.5215
Temperature	0.0105	0.4260
Assignment Time	0.0086	0.3441
Dispatch Time	0.0043	0.1703
Vendor Waiting Time	0.0042	0.1638
Vendor Acceptance Time	0.0037	0.1464
Monday	0.0025	0.0960
Lunchtime	0.0015	0.0556
Evening Hours	0.0012	0.0418
Friday	0.0011	0.0358
Wednesday	0.0010	0.0327
Saturday	0.0010	0.0320
Afternoon Snacks	0.0009	0.0298
Thursday	0.0003	0.0065
Late Evening Hours	0.0003	0.0031
Tuesday	0.0002	0.0014
Sunday	0.0002	0.0000

Table 3.5 : Normalised values banding and their corresponding linguistic terms

Normalised Values Banding	Linguistic Terms
0.9 - 1.0	Absolutely More Importance (AMI)
0.8 - 0.9	Very High Importance (VHI)
0.7 - 0.8	High Importance (HI)
0.6 - 0.7	Slightly More Importance (SMI)
0.5 - 0.6	Equally Importance (EI)
0.4 - 0.5	Slightly Low Importance (SLI)
0.3 - 0.4	Low Importance (LI)
0.2 - 0.3	Very Low Importance (VLI)
0.0 - 0.2	Absolutely Low Importance (ALI)

Table 3.6 : Spherical fuzzy numbers of criteria

Criteria	Spherical Fuzzy Numbers
Number of packages	(0.9, 0.1, 0.1)
Courier Transport Time	(0.9, 0.1, 0.1)
Order Preparation Time	(0.6, 0.4, 0.4)
Transaction Amount	(0.5, 0.5, 0.5)
Precipitation	(0.5, 0.5, 0.5)
Temperature	(0.4, 0.6, 0.4)
Assignment Time	(0.3, 0.7, 0.3)
Dispatch Time	(0.1, 0.9, 0.1)
Vendor Waiting Time	(0.1, 0.9, 0.1)
Vendor Acceptance Time	(0.1, 0.9, 0.1)
Monday	(0.1, 0.9, 0.1)
Lunchtime	(0.1, 0.9, 0.1)
Evening Hours	(0.1, 0.9, 0.1)
Friday	(0.1, 0.9, 0.1)
Wednesday	(0.1, 0.9, 0.1)
Saturday	(0.1, 0.9, 0.1)
Afternoon Snacks	(0.1, 0.9, 0.1)
Thursday	(0.1, 0.9, 0.1)
Late Evening Hours	(0.1, 0.9, 0.1)
Tuesday	(0.1, 0.9, 0.1)
Sunday	(0.1, 0.9, 0.1)

After scoring the alternatives and determining the criteria weights, the aggregated weighted spherical fuzzy decision matrix was created using equation 3 and the matrix is shown in Table 3.7.

Table 3.7 : Aggregated weighted spherical fuzzy decision matrix

Criteria	Multi-Package Delivery	Immediate Delivery
Number of packages	(0.72, 0.2227, 0.1039)	(0.45, 0.5074, 0.2291)
Courier Transport Time	(0.45, 0.5074, 0.2291)	(0.63, 0.3148, 0.1459)
Order Preparation Time	(0.18, 0.756, 0.4345)	(0.42, 0.4854, 0.3533)
Transaction Amount	(0.35, 0.5635, 0.4153)	(0.25, 0.6614, 0.559)
Precipitation	(0.25, 0.6614, 0.559)	(0.25, 0.6614, 0.559)
Temperature	(0.2, 0.7211, 0.5385)	(0.2, 0.7211, 0.5385)
Assignment Time	(0.15, 0.7858, 0.522)	(0.21, 0.7321, 0.346)
Dispatch Time	(0.03, 0.9503, 0.3119)	(0.07, 0.9095, 0.3055)
Vendor Waiting Time	(0.07, 0.9095, 0.3055)	(0.03, 0.9503, 0.3119)
Vendor Acceptance Time	(0.05, 0.926, 0.5025)	(0.07, 0.9095, 0.3055)
Monday	(0.05, 0.926, 0.5025)	(0.05, 0.926, 0.5025)
Lunchtime	(0.05, 0.926, 0.5025)	(0.07, 0.9095, 0.3055)
Evening Hours	(0.07, 0.9095, 0.3055)	(0.07, 0.9095, 0.3055)
Friday	(0.05, 0.926, 0.5025)	(0.07, 0.9095, 0.3055)
Wednesday	(0.05, 0.926, 0.5025)	(0.05, 0.926, 0.5025)
Saturday	(0.07, 0.9095, 0.3055)	(0.08, 0.9042, 0.2069)
Afternoon Snacks	(0.05, 0.926, 0.5025)	(0.07, 0.9095, 0.3055)
Thursday	(0.05, 0.926, 0.5025)	(0.05, 0.926, 0.5025)
Late Evening Hours	(0.03, 0.9503, 0.3119)	(0.07, 0.9095, 0.3055)
Tuesday	(0.05, 0.926, 0.5025)	(0.05, 0.926, 0.5025)
Sunday	(0.07, 0.9095, 0.3055)	(0.08, 0.9042, 0.2069)

For the defuzzification process, score function values were calculated by applying the equation 4 to the values in the aggregated weighted spherical fuzzy decision matrix.

The results of score function values are given in Table 3.8. For each criterion, the largest value indicates the positive ideal solution (PIS) and the smallest value indicates the negative ideal solution (NIS). The Score Function Values and Spherical Fuzzy numbers for each criterion obtained for PIS and NIS are presented in Table 3.9.

Table 3.8 : Score function values

Criteria	Multi-Package Delivery	Immediate Delivery
Number of packages	0,3655	-0,0287
Courier Transport Time	-0,0287	0,2058
Order Preparation Time	-0,0386	-0,013
Transaction Amount	-0,0177	0,085
Precipitation	0,085	0,085
Temperature	0,0812	0,0812
Assignment Time	0,0688	-0,1306
Dispatch Time	-0,3281	-0,3094
Vendor Waiting Time	-0,3094	-0,3281
Vendor Acceptance Time	0,0254	-0,3094
Monday	0,0254	0,0254
Lunchtime	0,0254	-0,3094
Evening Hours	-0,3094	-0,3094
Friday	0,0254	-0,3094
Wednesday	0,0254	0,0254
Saturday	-0,3094	-0,4701
Afternoon Snacks	0,0254	-0,3094
Thursday	0,0254	0,0254
Late Evening Hours	-0,3281	-0,3094
Tuesday	0,0254	0,0254
Sunday	-0,3094	-0,4701

Finally, spherical fuzzy numbers for PIS and NIS are determined as given in Table 3.9. After that, the distance between the PIS and NIS for each alternative was calculated using the normalised Euclidean distance through the equation 5 and equation 6. These distances are shown in Table 3.10. The maximum distance to Supherical Fuzzy Negative Ideal Solution (SFNIS) and the minimum distance to Supherical Fuzzy

Positive Ideal Solution (SFPIS) were obtained from Table 3.10. Closeness ratio was calculated using the equation 3.7 and the results are presented in Table 3.11.

Table 3.9 : PIS and NIS results

	Score Function Values for PIS	Spherical Fuzzy Numbers for PIS	Score Function Values for NIS	Spherical Fuzzy Numbers for NIS
Number of packages	0,3655	(0.72, 0.2227, 0.1039)	-0,0287	(0.45, 0.5074, 0.2291)
Courier Transport Time	0,2058	(0.63, 0.3148, 0.1459)	-0,0287	(0.45, 0.5074, 0.2291)
Order Preparation Time	-0,013	(0.42, 0.4854, 0.3533)	-0,0386	(0.18, 0.756, 0.4345)
Transaction Amount	0,085	(0.25, 0.6614, 0.559)	-0,0177	(0.35, 0.5635, 0.4153)
Precipitation	0,085	(0.25, 0.6614, 0.559)	0,085	(0.25, 0.6614, 0.559)
Temperature	0,0812	(0.2, 0.7211, 0.5385)	0,0812	(0.2, 0.7211, 0.5385)
Assignment Time	0,0688	(0.15, 0.7858, 0.522)	-0,1306	(0.21, 0.7321, 0.346)
Dispatch Time	-0,3094	(0.07, 0.9095, 0.3055)	-0,3281	(0.03, 0.9503, 0.3119)
Vendor Waiting Time	-0,3094	(0.07, 0.9095, 0.3055)	-0,3281	(0.03, 0.9503, 0.3119)
Vendor Acceptance Time	0,0254	(0.05, 0.926, 0.5025)	-0,3094	(0.07, 0.9095, 0.3055)
Monday	0,0254	(0.05, 0.926, 0.5025)	0,0254	(0.05, 0.926, 0.5025)
Lunchtime	0,0254	(0.05, 0.926, 0.5025)	-0,3094	(0.07, 0.9095, 0.3055)
Evening Hours	-0,3094	(0.07, 0.9095, 0.3055)	-0,3094	(0.07, 0.9095, 0.3055)
Friday	0,0254	(0.05, 0.926, 0.5025)	-0,3094	(0.07, 0.9095, 0.3055)
Wednesday	0,0254	(0.05, 0.926, 0.5025)	0,0254	(0.05, 0.926, 0.5025)
Saturday	-0,3094	(0.07, 0.9095, 0.3055)	-0,4701	(0.08, 0.9042, 0.2069)
Afternoon Snacks	0,0254	(0.05, 0.926, 0.5025)	-0,3094	(0.07, 0.9095, 0.3055)
Thursday	0,0254	(0.05, 0.926, 0.5025)	0,0254	(0.05, 0.926, 0.5025)
Late Evening Hours	-0,3094	(0.07, 0.9095, 0.3055)	-0,3281	(0.03, 0.9503, 0.3119)
Tuesday	0,0254	(0.05, 0.926, 0.5025)	0,0254	(0.05, 0.926, 0.5025)
Sunday	-0,3094	(0.07, 0.9095, 0.3055)	-0,4701	(0.08, 0.9042, 0.2069)

Table 3.10 : Distance between alternatives and SFPIS, alternatives and SFNIS

Alternatives	$D(X_i,X^*)$	$D(X_i,X^-)$
Multi-package Delivery	0.0788	0.0961
Immediate Delivery	0.0961	0.0788

Table 3.11 : Closeness ratio and ranking

Alternatives	Closeness ratio of each alternative	Ranking
Multi-package Delivery	0	1
Immediate Delivery	0.3996	2

The results of the analysis showed that for the specific month, multi-package delivery strategy is a more suitable option. This finding will guide operational planning such as warehouse organisation, updating the ordering system and staff training. It will increase the efficiency of the company by optimizing the delivery rates. This study has shown that the hybrid usage of SHAP values and Spherical Fuzzy TOPSIS method can be an effective tool in logistics decision making processes. In future studies, testing this approach in different sectors and with larger data sets will increase the generalisability of the method.

3.8 Discussion, Conclusion and Future Research

3.8.1 Conclusion and Discussion

The innovative hybrid model presented in this paper provides a versatile solution for optimizing delivery processes in the online retail industry. The developed system not only predicts order delivery success but also reveals the parameters affecting this success. In addition, the model integrated with the Spherical Fuzzy TOPSIS method provides monthly operational decision-making.

Thanks to this integrated approach, companies can make quick decisions by considering the probability of success in instant order planning and warehouse planning every month. This bi-directional advantage of the model is expected to provide significant benefits to companies in ecology (gasoline consumption and carbon dioxide emission reduction), cost management (optimization of fuel and storage costs) and customer satisfaction. The developed end-to-end innovative hybrid model also offers companies flexibility. Businesses will have the opportunity to make shorter-term and dynamic planning by reducing the planning period from monthly to weekly or even daily according to their needs.

This study's results significantly contribute to the development of data-driven decision-making processes in the online retail sector. The proposed model helps create a sustainable and competitive business model by increasing the efficiency of delivery processes while considering customer satisfaction.

3.8.2 Future Research

In the light of the results of this study, the following suggestions for future research are presented:

- Multiple Decision-Maker Approach: The current study adopted a single decision-maker approach in the Fuzzy TOPSIS application. In future studies, it is suggested that additional workshops be organized to obtain the opinions of different decision-makers and extend the model with a multiple decision-maker method.
- Expansion of the Decision-Making Model: The current model uses the Spherical Fuzzy TOPSIS decision-making model for monthly planning. In future studies, this model can be extended to allow daily or per-order decision-making.
- Expanding the Criteria Set: In addition to the existing criteria, new and more comprehensive criteria should be added, and the models should be strengthened accordingly.
- Large-Scale Testing and Optimisation: Testing the model's success in larger-scale online retail warehouses and optimizing the models according to the results obtained will increase its practical applicability.
- Proof of Efficiency of SHAP-SFS Transformation: Experimentally proving the difference between the transformation from SHAP values to Spherical Fuzzy Sets and the SFS approach generated directly by the decision maker would further demonstrate the methodology's effectiveness (Chebat et al, 2010).

These proposed research directions will contribute to further developing the presented model and its wider application in the online retail industry.

4. MULTI-CRITERIA FUZZY DECISION-MAKING TECHNIQUES WITH TRANSACTION COST ECONOMY PERSPECTIVES IN PRODUCT LAUNCH PROCESS

The transition from product development to productization is critical for data analytics solutions. In this thesis, the AI-based data models developed for use in the retail sector are conceptualized as digital products, since their deployment requires strategic planning, resource allocation, and integration into business processes. While many studies focus on modeling and development aspects, fewer studies address the strategic decision of whether products should be developed using internal resources or external partners. This section presents a study on transaction cost economics in the product launch process, particularly how multi-criteria fuzzy decision-making techniques can optimize this decision.Transaction cost theory, pioneered by Coase (1937) and later developed by Williamson (1975, 1981), provides a framework for understanding when organizations should perform activities internally and contract externally. This theory identifies three fundamental dimensions: uncertainty, asset specificity, and order frequency. Transaction cost economics is the difference between outsourcing and internal production costs.

This research addresses a significant gap in the literature by developing a decision-making method that will help companies determine the optimal resource allocation strategy (internal or external) during product or project development. The innovative contribution is creating a multi-criteria fuzzy decision-making framework that quantifies the impact of various cost factors on transaction costs, allowing companies to make more informed strategic decisions about resource allocation.

This section is based on the book chapter "Ozdemir, C., Cevik Onar, S., 2023: Multi-criteria Fuzzy Decision-Making Techniques with Transaction Cost Economy Theorem Perspectives in Product Launching Process. In: Calisir, F. (eds) Industrial Engineering in the Age of Business Intelligence. Lecture Notes in Management and Industrial Engineering. Springer, Cham."

The study used a comprehensive methodology that included expert input from product managers in large-scale companies with experience in product launches. Throughout this process, all costs arising in both internal production and outsourcing scenarios were identified, categorized, and prioritized using the Fuzzy Analytic Hierarchy Process (FAHP).

Research findings reveal that infrastructure and development costs, uncertainty, and time costs emerge as the most important factors affecting internal and external sourcing decisions. The study presents a systematic decision-making framework that can help companies optimize their product development strategies by determining the weights of these cost parameters that are of relative importance.

This research represents a significant step toward creating a decision model combining quantitative cost variables and qualitatively defined linguistic factors within the same framework. The insights obtained will help companies make more effective decisions in their product development processes and ultimately provide cost advantages in the retail sector.

4.1 Introduction

The fact that the transaction cost issue is an important paradigm has been provided by the studies and researches developed by Coase and then Williamson (Coase, 1937; Williamson O. E., 1975). There are 3 dimensions in the field of Transaction cost economies. These dimensions are uncertainty, asset specificity, and frequency of order (Williamson O. E., 1981). Transaction cost occurs when these dimensions encounter a certain behavior (bounded rationality) or opportunism (Englander, 1988). Transaction cost economy is defined as the total delta cost arising from the subtraction of the cost resulting from the outsourcing of a job from the cost of doing it with internal resources.

In the case of doing this job with external resources, the costs of researching the outsourced company that will do the job, making the contract, and monitoring this work arise. In case companies carry out these works with internal resources, the company has production costs. These costs and structures are affected by the 3 dimensions given above (Klaus, 1992).

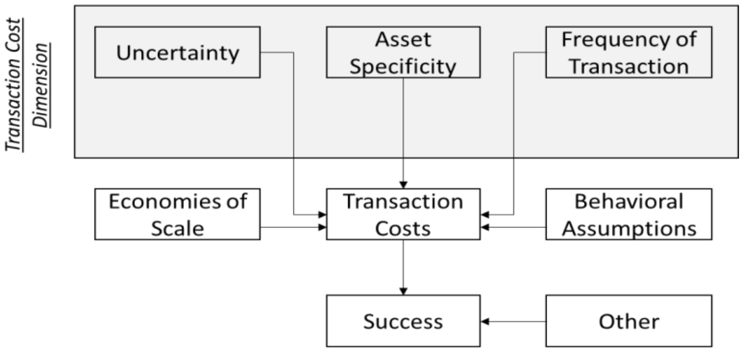

Figure 4.1 : Conceptual framework of transaction costs

One of the most important problems is that the effect of cost items included in the transaction cost structure on the total costs calculated is not in scalar form. This situation affects the decision-making mechanism of the companies. A new decision-making method is needed to determine the coefficient of cost items that affect the total transaction costs and to determine the strength of each cost item's effect. The main purpose of this article is to reveal this method. With this method to be revealed, it is to provide support to the company during the decision whether the company should do it with internal or external resources. This article reveals all the costs that the company has to bear in the case of internal production or outsourcing. These costs were added to the decision-making technique by categorizing them into 3 main dimensions. With this technique, companies will be able to easily see which decision to take in which situations. Within the scope of this article, some of the variables determined to be used in the decision-making model are known numerically, but some of them cannot be known and predicted during the productization process. For the purpose of this study, it also expresses linguistically those that cannot be defined by value. In addition, it is to create a decision model by using value-based variables and linguistic-based variables in the same model. In line with this publication, the prioritization and weighting of all the variables, which is the first stage of the planned study, will be done linguistically. In the next stage, the article will be expanded with the establishment of the planned decision-making model in line with this output.

4.2 Background

Many studies have been conducted in the field of transaction cost in the literature. In some of these studies, from the perspective of transaction cost economics, which decisions companies take in the event of a strategy or innovation shock have been examined. These investigations will show that the tendency to accept a higher transaction cost rate will be higher to quickly move to a new position in response to the shock. This situation will occur if the cost of moving late to the new position exceeds the contribution of early action (Argyres & Nickerson, 2019). In addition, according to the study, firms that are already repositioned are more likely to prefer to do these jobs with internal sources to reduce transaction costs as competitive pressure increases. From another point of view, it is stated that during the innovation shock, companies interested in innovation have a higher ability to imitate compared to other companies and can use outsourced resources to imitate this innovation quickly at the beginning (Bigelow & Nickerson, 2019). In this study, it has been observed that as the competition increases and the imitations increase, production returns through internal resources. In another study conducted in terms of transaction costs, it is revealed that retail companies are more motivated to turn to private branding instead of using the manufacturer's brand in product marketing (Chen S.-F. , 2010). It is stated that they use this situation to provide the commitment to the outsourcing company to do the work. Optimization and operation studies were also carried out in the area of transaction costs. Examples of these studies are the selection of the areas that bring the investment decision to the highest profitability and portfolio optimization area by using the transaction cost theory (Best & Hlouskova, 2005; Gerard & Alan, 1994). In addition, this theorem led by Coase and Williamson is also used in decision-making situations. These analyzes were provided by regression, chi-square analysis, OLS, and other statistical analyses.

Table 4.1 : Transaction cost literature review based on scopus library (method breakdown)

	Transaction cost		**Transaction Cost & Decision Making**		**Transaction Cost & Fuzzy**		**Transaction Cost & Fuzzy & Decision Making**	
Year	**Paper**	**Article**	**Paper**	**Article**	**Paper**	**Article**	**Paper**	**Article**
1950-1999	3.608	2.939	129	111	8	4	1	0
2000-2009	9.738	6.714	503	321	67	34	17	10
2011-2021	15.030	10.347	804	574	190	133	26	21
Total	**28.376**	**20.000**	**1.436**	**1.006**	**265**	**171**	**44**	**31**

As mentioned above, the studies in the transaction cost area are spread over many areas. Generally, it is aimed to find out which decision companies take when they need strategic change and the variables that affect this decision. In this direction, it was examined what is important when making decisions. It also examined the change in transaction costs before and after these decisions. In line with this information, it is also determined that which cost items affect the transaction cost and in which case they provide production with internal resources, in which case they operate with outsourcing.

The literature study in the field of transaction costs revealed the broad research perspective mentioned above. However, in these studies, there are very few studies involving a decision-making technique in line with the transaction cost perspective of whether the company should do a job with internal or external resources. In line with the research conducted within the scope of this paper, a decision-making technique that is created in line with the transaction cost dimensions (asset specificity, uncertainty, frequency, and behavioral assumption) will be formulated during the production and development of a strategic business. For this reason, it is necessary to use a fuzzy set because the effect of cost items in the determined dimensions is relative and not clear. The number of articles with both transaction cost and decision making and fuzzy is also limited in the literature. With this study, the transaction cost theorem and the multi-criteria fuzzy decision-making method were established. In this way, the research conducted will help the firm determine whether it should do the job with internal resources or outsourcing if it makes a strategic decision. In addition, this deficiency in the literature will be contributed.

4.3 Case Study

The main purpose of technology companies is to increase their own income and meet the needs of customers by developing new digital application products and projects. If the changing needs of customers are met with new products, the chance of achieving customer loyalty will increase. Higher-income and profitability are also obtained from loyal customers. In this direction, it is very important for technology companies to develop and launch products/projects that will meet the needs of customers and the market. In addition, technology companies must launch these products and projects at the right time and most optimal cost. It is one of the most important problems they need to solve, whether it should provide this product/project with internal or external resources. With the parameters obtained and determined within the scope of this article, a decision-making diagram will be established in which the source to proceed during the development of the digital product in the technology sector is determined. In this way, it will be possible to observe which variables affect the decision diagram and in which cases it is necessary to proceed with internal resources or external resources.

4.3.1 Data sources

One of the biggest problems in the process of product extraction in the technology sector is that some of the costs incurred in the development of the product or project are uncertain and not clearly determined. Although it is difficult to clarify the cost parameters within the scope of this study, a working group consisting of product manager experts working in large-scaled technology companies has been formed. In line with this working group, many product and project developments of product managers who have launched digital products have been examined in detail. The cost parameters that emerged during these developments were determined. Within the scope of this application, the required cost items and parameters in digital product and project development stages were determined by using the product development process capability of large-scale technology service providers. These technology companies include companies with more than 500 employees and companies that release digital products to meet customer needs. In addition, these companies actively use digital tools while producing products. In addition, the following cost parameter types are accessed, cleared, and corrected for analysis. Tables 4.2 show the variable categories and data types in detail.

Table 4.2 : Total cost variables (cost category breakdown)

HR	EC	SA	PA	UT
Human Resources	**Infrastructure & Development**	**Support Activities**	**Process & Approval**	**Uncertainty & Time**
Insource- Human Resources Personal Cost	Insource- Infrastructure & Maintenance Cost	Insource- Design & Market Research Cost	Insource- Approval Process Cost	Insource- Internal Uncertainty Cost
Insource- Project	Insource- Development & Test Cost	Insource- Sales &	Insource- Integration Cost	Insource- Time Cost

Table 4.2 (continued) : Total cost variables (cost category breakdown)

Management Cost		Marketing Cost		
Insource-Training Cost	Insource- Device & Supply Chain Cost	Insource-After Sales & Support Cost	Insource-Business Continuity Cost	Insource-Searching Cost
Insource-Team Building Cost	Outsource-Infrastructure & Maintenance Cost	Insource- Tax Cost	Outsource-Contracting Cost	Outsource-Searching Cost
Insource-Know How & IP Cost	Outsource-Development & Test Cost	Outsource-Design & Market Research Cost	Outsource-Monitoring Cost	Outsource-External Uncertainty
Outsource-Human Resources Personal Cost	Outsource-Device & Supply Chain Cost	Outsource-Sales & Marketing Cost	Outsource-Approval Process Cost	Outsource-Adaption Cost
Outsource-Project Management Cost		Outsource-After Sales & Support Cost	Outsource-Integration Cost	Outsource-Time Cost
Outsource-Training Cost		Outsource-Tax Cost	Outsource-Coordinating Cost	
Outsource-Team Building Cost			Outsource-Business Continuity Cost	
Outsource-Know How & IP Cost				

4.3.2 Data models

With digital transformation, the need to determine whether the planned products or projects should be made with internal resources or outsourced has increased. The importance of making this decision right has become even more crucial, as it is the goal of technology companies to produce the fastest and most cost-effective products that will meet the needs of the market and customer with this transformation. So far, companies have made this decision by looking at the total delta cost. However, if the importance of all cost items is taken equal, it is possible to make wrong decisions. For this reason, the cost items determined in the above section have been categorized and included in the total delta cost formula below.

$$Total\ Cost = \Delta(Technical\ Efficiency\ Cost) + \Delta(Agency\ Efficiency\ Cost) \quad (4.1)$$

$$\Delta(Technical\ Efficiency\ Cost) = \Delta(HR) + \Delta(ID) + \Delta(SA) \quad (4.2)$$

$$\Delta(Agency\ Efficiency\ Cost) = \Delta(PA) + \Delta(UT) \quad (4.3)$$

$$\Delta(HR) = IHRPC + IPMC + ITC + ITBC + IKIC - OHRPC - OPMC - OTC - OTBC - OKIC \quad (4.4)$$

$$\Delta(ID) = IIMC + IDTC + IDSC - OIMC - ODTC - ODSC \quad (4.5)$$

$$\Delta(SA) = IDMC + ISMC + IASC + ITAC - ODMC - OSMC - OASC - OTAC \quad (4.6)$$

$$\Delta(PA) = IAPC + IIC + IBCC - OCC - OMC - OAPC - OIC - OCOC - \text{OBCC} \quad (4.7)$$

$$\Delta(UT) = IIUC + ITIC - OSC - OEUC - OAC - OTIC \quad (4.8)$$

The effect of each of the determined cost items on the job output and the decision does not have the same weight. The fuzzy analytic hierarchy process (FAHP) has been applied in this study in order to learn the order of importance and weight of these cost items. The order of importance and weights of cost items were determined by the fuzzy analytical hierarchy process. In this study, the fuzzy analytical hierarchy process (FAHP) has been applied to find the importance and weight of cost items. The significance weight revealed by using FAHP will be added to the delta cost formula. The following steps have been applied to find and prioritize the weights of cost items with FAHP (Ozdemir et al., 2021).

The first step of FAHP is to give an important value to the features with respect to the Prioritization Scale table. After the prioritization process of features, the pairwise comparison matrix is established based on the a_{ij}. Moreover, Buckley's Fuzzy AHP Method was applied to find normalized weights of features (Buckley, 1985).

Table 4.3 : Prioritization scale

Value	Numbers (a_{ij})
Equal	(1,1,1)
Equal to Moderate	(1,2,3)
Moderate	(2,3,4)
Moderate to Strong	(3,4,5)
Strong	(4,5,6)

Table 4.3 (continued) : Prioritization scale

Strong to Very Strong	(5,6,7)
Very Strong	(6,7,8)
Very Strong to Extremely Strong	(7,8,9)
Extremely Strong	(8,9,9)

Step 1. Each element ($\breve{a}_{ij}$) of the pairwise comparison matrix $\tilde{A}$ is a fuzzy number corresponding to its prioritization scale (Buckley, 1985):

$$\tilde{A} = \begin{vmatrix} 1 & \breve{a}_{12} & \cdots & \breve{a}_{1n} \\ \breve{a}_{21} & 1 & \ldots & \breve{a}_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ \breve{a}_{n1} & \breve{a}_{n2} & \ldots & 1 \end{vmatrix} \tag{4.9}$$

$$\begin{gathered} \breve{a}_{12} = (a_{12l}, a_{12m}, a_{12u}) \\ \vdots \\ \breve{a}_{1n} = (a_{1nl}, a_{1nm}, a_{1nu}) \end{gathered} \tag{4.10}$$

Step 2. The consistency of each fuzzy comparison matrix is examined. Pairwise comparison values are defuzzified by the graded mean integration approach in order to check the consistency of the fuzzy pairwise comparison matrices. According to the graded mean integration approach, a triangular fuzzy number can be transformed into a crisp number by employing the equation below (Buckley, 1985):

$$A = \frac{l + 4m + u}{6} \tag{4.11}$$

Step 3. The fuzzy geometric mean for each row of matrices is computed to weigh the criteria and alternatives (Buckley, 1985).

$$\begin{gathered} a_{1l} = [1 \times a_{12l} \times \ldots \times a_{1nl}]^{1/n} \\ a_{2l} = [a_{21l} \times 1 \times \ldots \times a_{2nl}]^{1/n} \\ \cdots \\ a_{il} = [a_{n1l} \times a_{n2l} \times \ldots \times 1]^{1/n} \end{gathered} \tag{4.12}$$

$$b_{1l} = [1 \times b_{12l} \times \ldots \times b_{1nl}]^{1/n}$$
$$b_{2l} = [b_{21l} \times 1 \times \ldots \times b_{2nl}]^{1/n} \tag{4.13}$$
$$\ldots$$
$$b_{il} = [b_{n1l} \times b_{n2l} \times \ldots \times 1]^{1/n}$$

$$c_{1l} = [1 \times c_{12l} \times \ldots \times c_{1nl}]^{1/n}$$
$$c_{2l} = [c_{21l} \times 1 \times \ldots \times c_{2nl}]^{1/n} \tag{4.14}$$
$$\ldots$$
$$c_{il} = [c_{n1l} \times c_{n2l} \times \ldots \times 1]^{1/n}$$

Assume that the sums of the geometric mean values in the row are a_{1s} for lower parameters, a_{2s} for medium parameters, a_{1s} for upper parameters

$$\tilde{r}_{ij} = \begin{vmatrix} (\frac{a_{1l}}{a_{3s}}, \frac{b_{1m}}{a_{2s}}, \frac{c_{1u}}{a_{1s}}) \\ (\frac{a_{2l}}{a_{3s}}, \frac{b_{2m}}{a_{2s}}, \frac{c_{2u}}{a_{1s}}) \\ \vdots \\ (\frac{a_{il}}{a_{3s}}, \frac{b_{im}}{a_{2s}}, \frac{c_{iu}}{a_{1s}}) \end{vmatrix} \tag{4.15}$$

Step 4. The fuzzy weights and fuzzy performance scores are aggregated as follows:

$$\widetilde{U}_i = \sum_{\mathrm{j=1}}^{n} \widetilde{w}_j \tilde{r}_{ij}, \quad \forall i. \tag{4.16}$$

Step 5. The center of area method can be applied for defuzzification:

$$BNP_i = \frac{(u_i - l_i) + (m_i - l_i)}{3} + l_i, \quad \forall i \tag{4.17}$$

Step 6. The best alternative is determined as in the Fuzzy AHP

4.3.3 Data models evaluation

In the data analysis used in determining the weights of cost items, the steps (1) - (6) stated above were followed. Prioritization scale was applied for the Variable Category, and the subcategories were evaluated with FAHP. Table 4.4 shows the prioritization matrix of the main cost variables of transaction cost in line with FAHP analysis.

Table 4.4 : Scale table of main cost

	HR	ID	SA	PA	UT
HR	(1,1,1)	$(\frac{1}{6},\frac{1}{5},\frac{1}{4})$	(2,3,4)	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$
ID	(4,5,6)	(1,1,1)	(6,7,8)	(4,5,6)	(1,2,3)
SA	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	$(\frac{1}{8},\frac{1}{7},\frac{1}{6})$	(1,1,1)	$(\frac{1}{6},\frac{1}{5},\frac{1}{4})$	$(\frac{1}{8},\frac{1}{7},\frac{1}{6})$
PA	(1,2,3)	$(\frac{1}{6},\frac{1}{5},\frac{1}{4})$	(4,5,6)	(1,1,1)	$(\frac{1}{6},\frac{1}{5},\frac{1}{4})$
UT	(2,3,4)	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$	(6,7,8)	(4,5,6)	(1,1,1)

Transaction Cost main cost items are broken down into 6 categories. For the subgroups of these 6 categories, the FAHP prioritization matrices mentioned above were created separately. In order to take up less space in this article, only the prioritization table of the subgroups of the first categories is shown in Table 4.5.

Table 4.5 : Scale table infrastructure and development (ID)

	IIMC	IDTC	IDSC	OIMC	ODTC	ODSC
IIMC	(1,1,1)	$(\frac{1}{5},\frac{1}{4},\frac{1}{3})$	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	(2,3,4)	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$
IDTC	(3,4,5)	(1,1,1)	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	(3,4,5)	(1,2,3)	(1,2,3)
IDSC	(2,3,4)	(2,3,4)	(1,1,1)	(2,3,4)	(2,3,4)	(1,2,3)
OIMC	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	$(\frac{1}{5},\frac{1}{4},\frac{1}{3})$	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	(1,1,1)	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$	$(\frac{1}{6},\frac{1}{5},\frac{1}{4})$
ODTC	(2,3,4)	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$	$(\frac{1}{4},\frac{1}{3},\frac{1}{2})$	(1,2,3)	(1,1,1)	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$
ODSC	(1,2,3)	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$	$(\frac{1}{3},\frac{1}{2},\frac{1}{1})$	(4,5,6)	(1,2,3)	(1,1,1)

After Pairwise Comparison Matrices were generated for corresponding Categories, normalized weights were calculated. The following weights and normalized weights for each category are in ascending order:

Table 4.6 : Weights of category of main variables

Feature	Weights	Normalized Weights
HR	0.10	0.09
ID	0.47	0.44
SA	0.04	0.04
PA	0.12	0.12
UT	0.34	0.31

Table 4.7 : Weights of category of ID variables

Feature	Weights	Normalized Weights
IIMC	0.10	0.08
IDTC	0.26	0.23
IDSC	0.37	0.31
OIMC	0.06	0.05
ODTC	0.15	0.13
ODSC	0.22	0.19

The consistency of the prioritization matrix created within the scope of FAHP was examined in accordance with step 2. Since the consistency rate of the given prioritization values is less than 10% for each matrix, it has been shown that the study is consistent.

4.4 Conclusion and Future Research

In line with the FAHP study, the importance of all group and subgroup variables has been determined linguistically in terms of cost. In this direction, when we examine the main cost group results, the Human Resources cost weight is 9%, the Infrastructure and Development cost weight is 44%, the Support Activities cost weight is 4%, the Process and Approval cost weight is 12%, and the Uncertainty and Time cost weight is 31%. In this direction, as a result of the points given by the experts, the cost of infrastructure and development and the cost of uncertainty and time became the most important costs. While the cost of infrastructure and development was the most important cost among technical efficiency, the cost of uncertainty and time was the most important cost among agency efficiency. In addition, the following results arose in the subgroups of infrastructure and development cost, which are examined in more detail within the scope of the study. Insource device and supply chain costs were the most important cost with a 31% weight rate. This cost was followed by Insource Development and Test costs, with a weight rate of 23%. On the outsourcing side, the device and supply chain cost have the highest importance with a 19% weight rate, while the development and testing cost has the second-highest importance with a 13% weight rate.

For both outsource and insource, the cost of infrastructure and maintenance is less important. This study has been graded in detail for all subgroups, and the importance coefficient has been revealed. In line with the importance coefficients here, it will be possible to calculate the total transaction cost more effectively by using the formulas between 4.1 and 4.8.

In line with all these studies, the order of priority and importance weight values of both main cost variables and subgroup variables were found among themselves. The aim of this study is to generate the groundwork for the decision-making structure that will be created by using both the variables defined by value and the linguistically defined variables in the same model. Moreover, this publication is to be the initial study of this decision-making structure. The paper will be expanded in line with the outputs from here. In line with this paper, we will create a decision model by using both value-based variables and linguistically -defined variables in the same model in the next stage. With the studies to be developed and detailed, it is planned that companies will be able to decide more clearly with which source they should use while developing the product.

5. CONCLUSION AND RECOMMENDATIONS

This thesis has comprehensively examined the development and implementation of data-based cost advantage models in the retail sector, covering physical and online retail areas. Through three interconnected studies, we have demonstrated how innovative applications of data analytics, fuzzy logic, and decision-making techniques can provide significant cost advantages in different retail contexts. The research offers theoretical contributions to the academic literature and practical implications for industry practitioners aiming to optimize their operations and decision-making processes via technology companies.

5.1 Synthesis of Research Findings

The first study presented a new hybrid fuzzy prediction method for predicting shopping mall visitor numbers. Combining machine learning algorithms with fuzzy multi-criteria decision-making techniques ensured high prediction accuracy of visitor numbers for three shopping malls in Istanbul. This study's original value lies in its approach of determining the features to be used in prediction models with fuzzy multi-criteria decision-making techniques and weighting these features with machine learning algorithms. The hybrid approach is more explainable than traditional forecasting models and also increases forecast accuracy.

The effects of various environmental factors such as weather data, financial market data, traffic density and internet trends on the number of visitors to shopping malls were analyzed in this study. According to the results of the analyses conducted with the Fuzzy Analytical Hierarchy Process (FAHP), it was revealed that internet trends, weather conditions and daily average traffic density factors were the most effective factors in predicting the number of visitors. In light of these findings, it is possible to create a more effective roadmap for shopping mall managers in determining their marketing and operational strategies. For example, by predicting the change in visitor numbers according to these factors, advantages can be obtained in the costs to be spent on daily personnel, energy consumption and promotional activities.

Another important finding of the study is testing the developed model's applicability to different shopping malls and obtaining high success rates. This result demonstrates the model's generalizability and practical applicability. The detailed performance metrics (adjusted R^2, success rate, VIF, Durbin-Watson) used to evaluate the model's accuracy support its statistical robustness.

The second study presented an innovative hybrid approach to delivery optimization in online retail by combining machine learning prediction capabilities with Spherical Fuzzy Sets and the TOPSIS decision-making method. This study's most original contribution is its methodology of transforming SHAP (Shapley Additive exPlanations) values obtained from machine learning models into Spherical Fuzzy Sets. This transformation establishes a bridge between model outputs' explainability and uncertainty management in decision-making.

The study developed a Stochastic Gradient Boosting model to predict the probability of successful delivery of orders using online retail delivery data. As a result of the SHAP analysis, the most important factors affecting delivery success were determined as the number of packages, courier transportation time, transaction amount and order preparation time. The importance levels of these factors were converted into spherical fuzzy sets and two different scenarios were evaluated, both an efficiency-focused multi-package strategy and a customer satisfaction-focused instant delivery strategy.

As a result of the analysis conducted with Spherical Fuzzy TOPSIS, it was determined that the multi-pack strategy is more suitable for a particular month. This finding provides a valuable framework for online retail industry actors to optimize their strategies according to different conditions and constraints. The decision model aims to reduce delivery costs, increase customer satisfaction and support environmental sustainability.

The third study addressed a critical gap in the literature on resource allocation decisions in the product development process. Applying FAHP processes to transaction cost economics, this research determined the cost factors that would guide companies' decision-making processes regarding the use of internal or external resources in the resource allocation process and then completed the critical prioritizations. A decision model was developed by restructuring frequently encountered problems in the transaction cost theory literature, such as uncertainty and transaction frequency dimensions, with a fuzzy logic perspective.

Research has shown that infrastructure and development costs (44%) and uncertainty and time costs (31%) are determined as the most effective factors in the decision-making mechanism in the resource allocation process. These two factors are followed by process and approval costs (12%), human resources costs (9%) and support activities costs (4%). In addition, the sub-components that constitute these cost factors are analyzed in detail in the study and the importance weight of each sub-component is determined.

These findings provide a systematic framework that will guide companies to effectively manage resource allocation in their product development processes. Determining the importance of cost factors on a project-by-project basis supports more conscious and cost-effective decisions. While proceeding with internal resources in projects where uncertainty and time costs are high may be more appropriate in terms of cost optimization, external resources may be more efficient in projects where infrastructure and development costs are high.

The common feature of all three studies is that they develop hybrid methodologies by integrating data sets created with real sector data with fuzzy logic approaches. This combination offers high potential for providing cost advantage in the retail sector by increasing forecast accuracy and the effectiveness of decision-making processes.

5.1.1 Theoretical contributions

This thesis makes several important theoretical contributions to the fields of retail analytics, decision science, and operations management:

Firstly, it advances the methodological integration of machine learning with fuzzy decision-making techniques. Traditionally, machine learning algorithms are used separately in prediction and classification problems, while fuzzy logic approaches are used in decision problems involving uncertainty. This thesis demonstrates that combining the strengths of these two approaches makes it possible to develop more comprehensive and robust models. In particular, the predictive power of machine learning models and the uncertainty management capabilities of fuzzy logic approaches provide significant advantages in complex and dynamic environments such as retail.

This integration is clearly seen in the shopping mall visitor prediction and delivery optimization models. In the shopping mall model, machine learning algorithms increase prediction accuracy, while fuzzy multi-criteria decision-making techniques optimize the selection and weighting of model parameters. In the delivery model, SHAP values of machine learning predictions are transformed into spherical fuzzy sets and included in the decision-making process. This approach increases both the prediction performance and the explainability of the models.

Secondly, the thesis extends transaction cost theory by providing a quantitative framework for evaluating make-or-buy decisions in product development. Transaction cost theory explains how firms decide which activities to internalize and which to procure from the market. However, applying the theory often remains at the conceptual level or is used as an ex-post analysis. This thesis presents a more structured and quantitative decision model by combining transaction cost theory with fuzzy logic approaches. In particular, this thesis divides the three fundamental dimensions of transaction cost theory, uncertainty, asset specificity, and transaction frequency, into more detailed and measurable subcomponents. Determining the importance level of subcomponents with the fuzzy analytical hierarchy process makes the theory more concrete and applicable.

Thirdly, the research brings a new approach to modeling uncertainty in decision-making processes by transforming SHAP values into Spherical Fuzzy Sets. The use of SHAP values to explain the predictions produced by machine learning models is becoming increasingly common. However, the literature on the direct inclusion of these values in decision-making processes is limited. This thesis fills this gap in the literature. This transformation bridges the gap between uncertainty management in multi-criteria decision-making processes and estimation models created with machine learning models. It enables better understanding of machine learning estimations in decision-making processes and more effective decisions. In addition, SHAP values converted into fuzzy sets provide a more objective and data-driven alternative to traditional methods.

Fourthly, this thesis demonstrates the effectiveness of hybrid fuzzy prediction methods and contributes to the literature in the field of customer prediction and location analysis in the retail industry. The shopping mall visitor prediction model, which combines environmental, financial and behavioral variables, provides higher accuracy than traditional models.Demonstrating the applicability of the model in different shopping malls is important to increase the value of the approach. This finding shows that similar hybrid approaches can provide cost advantages in different locations in the retail sector.

Finally, this thesis contributes to the logistics by demonstrating the application of Spherical Fuzzy Sets in delivery operations in retail sector. This hybrid approach allows for evaluation by considering multiple and conflicting criteria such as operational efficiency and customer satisfaction when creating delivery strategies. In addition, the framework created for monthly delivery planning has a flexible structure, which enables rapid adaptation to dynamic market conditions and makes an important contribution to the retail logistics literature.

5.1.2 Practical contributions

This thesis provides theoretical contributions as well as many implications for retail managers, companies and the logistics sector.

In order to gain cost advantage in the competitive environment of shopping malls and physical retailers, it is critical to accurately predict customer traffic. The hybrid fuzzy model makes the impact of environmental factors such as weather, internet trends, financial markets on visitor numbers more understandable for decision makers and enables them to make more effective predictions.

These predictions allow for more effective resource allocation in operational areas such as personnel planning, energy consumption management, security services, cleaning, and maintenance services. For example, labor costs can be optimized by reducing the number of personnel on days when low visitor numbers are expected and increasing the number of personnel on days when high visitor numbers are expected. Similarly, energy consumption, security, and cleaning services can also be adjusted according to visitor predictions.

Additionally, visitor predictions offer valuable insights for planning marketing and promotional activities. Shopping mall managers can organize special events, discounts, or promotions during periods when low visitor numbers are expected, increasing traffic and preventing revenue loss. During periods when high visitor numbers are expected, they can implement targeted marketing activities and sales optimization strategies to leverage the existing traffic best.

The model can also optimize the tenant mix in shopping malls and rent negotiations. Visitor predictions can be used to analyze the performance of different tenant categories and shape the future tenant strategy of the shopping mall. For example, understanding which types of stores attract more customers in a certain period or weather condition provides valuable information for tenant mix optimization and rent pricing.

For online retailers and logistics providers, the hybrid approach to delivery optimization offers a powerful tool for balancing efficiency with customer satisfaction. With the rapid growth of e-commerce, last-mile delivery optimization has become a strategic priority for online retailers. Delivery costs are one of the most significant items of e-commerce operations and are a critical factor affecting customer satisfaction and loyalty.

The model analyzes factors affecting delivery success with machine learning and integrates these results with the spherical fuzzy TOPSIS method, allowing for evaluating different delivery strategies. Online retailers can use this model to determine which delivery strategy (multi-package or instant delivery) is more suitable for a specific period or region, optimize delivery routes, and reduce delivery failure rates.

In particular, the model's demonstration of the impact of factors such as the number of packages, courier carrying time, order preparation time, and transaction amount on delivery success provides valuable insights for online retailers to improve their operational processes. For example, in cases where order preparation time significantly affects delivery success, strategies such as increasing the efficiency of warehouse operations, developing a more effective pickup process, or employing more warehouse personnel can be implemented.

Additionally, since the model allows for evaluating the impact of different delivery scenarios on multiple criteria, online retailers can balance potentially conflicting goals such as operational efficiency, customer satisfaction, cost optimization, and environmental sustainability. For example, in a specific period or region, a multi-package delivery strategy may increase operational efficiency and reduce delivery costs but adversely affect customer satisfaction. The model provides a data-driven framework to optimize this balance.

This approach is valuable, especially for online retailers offering fast delivery, same-day delivery, or delivery in selected time intervals. The model can help optimize decisions on when and where to offer these services, plan delivery capacity more effectively, and better manage customer expectations.

The transaction cost decision framework offers a systematic approach to the make-or-buy decision in product development for companies and product managers. With the acceleration of digitalization and technological developments, companies are increasingly forced to develop new products and projects. In this process, the decision of which activities to perform with internal resources and which activities to perform with external resources is critically important in terms of cost efficiency and competitive advantage.

The framework allows for systematically evaluating and prioritizing the cost factors affecting this decision. Companies can make more informed resource allocation decisions by analyzing main cost categories such as infrastructure and development costs, uncertainty and time costs, process and approval costs, human resources costs, support activities costs, and their sub-components.

In particular, internal resource use may be more appropriate for projects with high uncertainty and time costs (for example, new and untested technologies, rapidly changing market conditions, or uncertain customer requirements), as internal resource use provides more control, flexibility, and rapid adaptation. On the other hand, external resource use may be more cost-effective for projects with high infrastructure and development costs (for example, large-scale infrastructure investments, projects requiring specialized hardware, or developments requiring rare expertise).

Additionally, framework allows for the evaluation of hybrid strategies. Companies can adopt different resource allocation strategies for different project components, developing components related to strategic importance or core competencies internally while procuring more standard or low-strategy-importance components externally.

This framework offers a valuable tool, especially for companies in digital transformation, technology startups, and businesses developing innovative products. By making resource allocation decisions more systematic and data-driven, companies can optimize development costs, shorten time to market, and gain a competitive advantage.

In all three areas, this research demonstrates the value of data-driven decision-making approaches that can cope with the complexity and uncertainty natural in retail operations. The thesis bridges the gap between advanced analytical techniques and real-world retail challenges by providing practical frameworks and methodologies that can be implemented with existing data and technologies. Finally, from the applicable perspective, these detailed applications can be integrated into other companies and sectors with minimal effort, thanks to the flexibility of the hybrid methodology

5.2 Limitations and Future Research Directions

While this thesis makes significant contributions to both theory and practice, several limitations must be acknowledged, each pointing to promising directions for future research:

First, the shopping mall visitor prediction model was tested in three shopping malls in a specific geographic region (the Anatolian side of Istanbul) and for a limited period (June 2018 - January 2019). This limited sample restricts the model's generalizability in different geographic regions, shopping mall formats, and extended periods. Additionally, the data sources and variables used in the study may not cover all factors affecting shopping mall visitor behaviors. Future research should expand the geographic scope and temporal duration to test the model's generalizability across different retail contexts, cultural environments, and economic conditions. The model's performance should be examined for different shopping mall formats (traditional malls, open-air malls, outlet centers, etc.), different retail categories (fashion, food and beverage, entertainment, etc.), and different consumer segments.

Including more real-time data sources and dynamic variables could further enhance the model's predictive power. Integrating new data sources such as social media, mobile app usage, customer demographics, and micro-location data into the model may enable a more comprehensive understanding and prediction of visitor behaviors.

Furthermore, similar hybrid approaches can be developed to predict more granular performance metrics beyond visitor numbers, such as spending per visitor, visit duration, in-store circulation patterns, and category-based conversion rates. Such models allow shopping mall managers to develop more detailed and targeted optimization strategies.

Second, the delivery optimization model was developed using a specific dataset from the online grocery sector. This focus may limit the model's applicability in different e-commerce categories and delivery scenarios. Additionally, the model evaluates only two delivery strategies (multi-package and instant delivery) and does not cover more complex and hybrid strategies.

Future research can apply this approach to other e-commerce categories (fashion, electronics, furniture, etc.) with different logistics requirements and customer expectations. Examining how factors affecting delivery success change for each category will increase the generalizability and practical value of the model. For example, product value and security may be more important factors in electronic products, while size and installation requirements may be more important factors in furniture products.

Additionally, the model can be expanded to evaluate more complex delivery scenarios and hybrid strategies. For example, similar analysis and optimization approaches can be developed for alternative delivery models such as click-and-collect, pickup points, time-window delivery, delivery boxes, and carbon-neutral delivery.

Furthermore, including environmental sustainability metrics in the decision framework would align with increasing concerns about the ecological impact of last-mile delivery operations. Integrating environmental factors such as carbon emissions, energy consumption, packaging waste, and traffic congestion into delivery decisions will contribute to developing more sustainable e-commerce operations.

Third, the transaction cost decision framework primarily focuses on this concept through technology companies and product development. This focus may limit the framework's applicability in other industries and product types. Additionally, the framework is currently a conceptual model and has not been sufficiently validated with real-world applications.

Future research can extend this framework to different industries and product types. Examining how the relative importance of cost factors changes for each industry and product type will increase the framework's generalizability and applicability. For example, in physical product development, infrastructure and development costs and human resources costs may be more important, while in service development, uncertainty and process costs may be more critical.

Additionally, real-world case studies and longitudinal research should be conducted to increase the validity and reliability of the framework. Such studies will allow for analyzing the long-term outcomes and performance impacts of resource allocation decisions and will test the framework's predictive power. For example, critical performance indicators such as cost efficiency, time to market, product quality, and customer satisfaction can be monitored for companies making decisions based on the framework.

The transaction cost framework should also be continuously updated in light of technological developments and changes in business models. For example, new trends such as generative ai, agentic ai, cloud computing, open-source software and platforms, customer-centered development methodologies, and digital ecosystems affect resource allocation decisions and transaction costs. The framework should evolve to reflect such trends and changes.

Fourth, while this research demonstrates the value of integrating machine learning with fuzzy decision-making techniques, there is room for further methodological improvements and comparisons with alternative approaches. The current study used specific machine learning algorithms (regression, stochastic gradient boosting, etc.) and specific fuzzy set extensions (FAHP, spherical fuzzy sets, etc.). However, whether these choices are optimal or other combinations would provide better performance has not been comprehensively examined.

Future studies can systematically compare machine learning algorithms, fuzzy set extensions, and decision-making methods to determine the most effective combinations for retail challenges. Different combinations of ai and machine learning approaches such as generative ai, agentic ai, deep learning, reinforcement learning, genetic algorithms, and fuzzy set extensions such as intuitionistic fuzzy sets, type-2 fuzzy sets, and z-sets can be examined.

Additionally, more research should be conducted in areas such as model validation, uncertainty estimation, sensitivity analysis, and interpretability to increase the performance and reliability of hybrid approaches. In particular, the relationship between uncertainty estimates of machine learning models and uncertainty representations of fuzzy sets should be examined in more detail.

Finally, as retail continues to evolve towards multi-channel models that blur the distinction between physical and online shopping, future research should explore integrated data-based models that span these areas. Current studies generally develop separate models for physical retail and online retail, and the integration and synergies of these two areas are not sufficiently examined.

Developing unified frameworks that optimize operations across channels while providing seamless customer experiences represents an important frontier for retail analytics research. For example, hybrid models where physical stores are also used as micro-distribution centers and delivery points will require data analytical approaches to optimize both physical and online retail operations.

Such unified frameworks should analyze and optimize customer behaviors and interactions at different stages of the customer journey (research, evaluation, purchase, delivery, return, etc.) and in different channels (physical store, website, mobile app, social media, etc.). This approach will enable retailers to develop and implement truly omni-channel strategies.

REFERENCES

Adu, J., Owusu, M. F., Martin-Yeboah, E., Pino Gavidia, L. A., & Gyamfi, S. (2022). A discussion of some controversies in mixed methods research for emerging researchers. *Methodological Innovations, 15*(3), 321-330.

Ak, M. F., & Demir, E. (2024). A novel qualitative risk assessment using the interval-valued spherical fuzzy extension of TOPSIS method: a case study in rail transit systems. *Neural Computing and Applications, 36*, 5109–5132.

Akram, M., Kahraman, C., & Zahid, K. (2021). Extension of TOPSIS model to the decision-making under complex spherical fuzzy information. *Soft Computing, 25*, 10771–10795.

Alonso, S., Cabrerizo, F. J., Chiclana, F., Herrera, F., & Herrera-Viedma, E. (2007). An interactive decision support system based on consistency criteria. *Journal of Mult.-Valued Logic & Soft Computing, 14*, 371-385.

Alshammari, I., Parimala, M., Ozel, C., & Riaz, M. (2022). Spherical linear diophantine fuzzy TOPSIS algorithm for green supply chain management system. *Journal of Function Spaces*.

Argyres, N., & Nickerson, J. (2019). Strategic responses to shocks: Comparative adjustment costs, transaction costs, and opportunity costs. *Strategic Management Journal, 40*(3), 357-376.

Ashraf, S., Abdullah, S., Aslam, M., Qiyas, M., & Kutbi, M. (2019). Spherical fuzzy sets and its representation of spherical fuzzy t-norms and t-conorms . *Journal of Intelligent & Fuzzy Systems, 36*(6), 6089–6102.

Atanassov, K. (1999). *Intuitionistic Fuzzy Sets.* Heidelberg: Springer.

Barbara, G. T., & Linda, S. F. (2013). *Using multivariate statistics.* Boston: Pearson.

Best, M. J., & Hlouskova, J. (2005). An algorithm for portfolio optimization with transaction costs. *Management Science, 51*(11), 1676-1688.

Bigelow, L., & Nickerson, J. (2019). When and how to shift gears: Dynamic trade-offs among adjustment, opportunity, and transaction costs in response to an innovation shock. *Strategic Management Journal, 40*(3), 377-407.

Boran, F. E., Genç, S., Mustafa, K., & Akay, D. (2009). A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. *Expert Systems with Applications, 36*(8), 11363-11368.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing, 93*(1), 79-95.

Buckley, J. J. (1985). Fuzzy hierarchical analysis. *Fuzzy Sets Systems,, 17*(1), 233–247.

Buyukozkan, G., & Feyzioglu, O. (2007). Evaluation of suppliers' environmental management performances by a fuzzy compromise ranking technique. *Journal of Mult.-Valued Logic & Soft Computing, 14,* 309-323.

Cevik Onar, S., & Yasin Ates, N. (2007). A fuzzy model for operational supply chain optimization problems. *Journal. of Mult.-Valued Logic & Soft Computing, 14,* 355-370.

Chang, D. (1992). Extent analysis and synthetic decision,. *Optimization Techniques and Applications, 1*, 352.

Chebat, J., Sirgy, M., & Grzeskowiak, S. (2010). How can shopping mall management best capture mall image. *Journal of Business Research, 63*(7), 735-740 .

Chen, C. T. (2000). Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets Syst, 114*(1), 1–9.

Chen, S.-F. (2010). Transaction cost implication of private branding and empirical evidence. *Strategic Management Journal., 31*(4), 371-389.

Cheng, E. W., Li, H., & Yu, L. (2005). The analytic network process (ANP) approach to location selection: A shopping mall illustration. . *Construction Innovation, 5*(2), 83-97.

Coase, R. H. (1937). The Nature of the Firm," Economica. *Economica, 4*, 368-405.

Coban, V., & Cevik Onar, S. (2017). Analysis of solar energy generation capacity using hesitant fuzzy cognitive maps. *International Journal of Computational Intelligence Systems, 10*(1), 1149-1167.

Coban, V., & Cevik Onar, S. (2018). Pythagorean fuzzy engineering economic analysis of solar power plants. *Soft Computing, 22*(15), 5007-5020.

Cote, J.-F., Queiroz, T., Gallesi, F., & Iori, M. (2018). Optimization methods for the same-day delivery problem. *Advances in Optimization and Decision Science for Society, Services and Enterprises* (s. 335–349). içinde

Duman, G. M., Tozanli, O., Kongar, E., & Gupta, S. M. (2017). A holistic approach for performance evaluation using quantitative and qualitative data: A food industry case study. *Expert Systems With Applications, 81*, 410-422.

Ehrler, V., Schoeder, D., & Seidel, S. (2019). Challenges and perspectives for the use of electric vehicles for last mile logistics of grocery e-commerce – findings from case studies in germany. *Research in Transportation Economics 8*.

Eiselt, H., & Marianov, V. (2011). *Foundations of Location Analysis.* London: Springer.

Eiselt, H., & Marianov, V. (2015). *Applications of Location Analysis.* London: Springer.

El-Adly, M. (2007). Shopping malls attractiveness: a segmentation approach. *International Journal of Retail & Distribution Management , 35*(11), 936–950.

Englander, E. J. (1988). Technology and Oliver Williamson's transaction cost economics. *Economic Behavior and Organization, 10*, 339-353.

Fawcett, T. (2006). Introduction to roc analysis. *Pattern Recognition Letters, 27*, 861–874.

Fayyad, U., Piatetsky-Shapiro, G., & P., S. (1996). From data mining to knowledge discovery in databases. . *AI Magazine , 17*(3), 37-54.

Field, A. (2009). *Discovering Statistics using SPSS.* London: Sage.

Friedman, J. (2002). Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, 367-378.

Gauri, D. K., Jindal, R. P., Ratchford, B., Fox, E., Bhatnagar, A., Pandey, A., . . . Howerton, E. (2021). Evolution of retail formats: Past, present, and future. *Journal of Retailing, 97*(1), 42-61.

Gauvreau, C. A. (2018). Rehabilitation strategies enhancing participation in shopping malls for persons living with a disability. . *Disability and Rehabilitation , 40*(8), 917–925.

Gedik, Y. (2024). Perakende sektöründe yapay zekâ avantajları, zorlukları ve stratejileri üzerine kavramsal bir değerlendirme. *Kastamonu İnsan ve Toplum Dergisi – KİTOD, 2*(4), 345-368.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th b.).* Boston: MA: Allyn & Bacon.

Gerard, G., & Alan, J. (1994). Investment strategies under transaction costs: The finite horizon case. *Management Science, 40*(3), 385-404.

Gevaers, R., Voorde, E., & Vanelslander, T. (2011). Characteristics and typology of last-mile logistics from an innovation perspective in an urban context, in: city distrib. . *Urban Freight Transport: Multiple Perspectives.* içinde Edward Elgar Publishing.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis.* New York: Macmillan.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine Learning, 45*, 171–186.

Hedhli, K. E., Chebat, J.-C., & Sirgy, M. J. (2013). Shopping well-being at the mall: Construct, antecedents, and consequences. . *Journal of Business Research , 66*(7), 856–863 .

Heldt, B., Matteis, T., Schmidt, A., & Heinrichs, M. (2019). Cool but dirty food? – Estimating the impact of grocery home delivery on transport and CO2 emissions including cooling. *Research in Transportation Economics*.

Huotari, V. (2015). *Depth camera based customer behaviour analysis for retail: Master's thesis.* Oulu : University of Oulu .

Hwang, C., & Yoon, K. (1981). Multiple Attribute Decision Making. Lecture Notes in Economics and Mathematical Systems. In *Methods for Multiple Attribute Decision Making.* Berlin, Heidelberg: Springer.

Jacobs, B., Donkers, B., & Fok, D. (2016). Model-based purchase predictions for large assortments. *Marketing Science*.

Jaller, M., & Otay, I. (2021). Evaluating Sustainable Vehicle Technologies for Freight Transportation Using Spherical Fuzzy AHP and TOPSIS. *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions. INFUS 2020. Advances in Intelligent Systems and Computing,.* içinde Springer, Cham.

Jiang, H., & Xinhui, R. (2020). Comparative analysis of drones and riders in on-demand meal delivery based on prospect theory. *Discrete Dynamics in Nature and Society*, 1–13.

Kahraman, C. (2008). Fuzzy decision making and soft computing applications - Introduction. *Journal of Mult.-Valued Logic & Soft Computing, 14*.

Kahraman, C., & Kutlu Gundogdu, F. (2018). From 1D to 3D Membership:Spherical fuzzy sets. *BOS / SOR 2018, Polish Operational and Systems Research Society, September 24th – 26th.* Palais Staszic, Warsaw, Poland.

Kahraman, C., & Kutlu Gundogdu, F. (2020). *Decision making with spherical fuzzy sets theory and applications.* Springer.

Kahraman, C., Ates, N., Cevik, S., & Gülbay, M. (2007). Fuzzy multi-attribute cost–benefit analysis of e-services: Research Articles. *International Journal of Intelligent Systems, 22*(5), 547-565.

Kahraman, C., Cebeci, U., & Ruan, D. (2004). Multi-attribute comparison of catering service companies using fuzzy AHP: The case of Turkey. *International Journal of Production Economics, 87*, 171-184.

Kahraman, C., Cevik Onar, S., & Basar, O. (2015). Engineering economic analyses using intuitionistic and hesitant fuzzy sets. *Journal of Intelligent & Fuzzy Systems, 29*(23), 1151-1168.

Kahraman, C., Kutlu Gundogdu, F., Cevik Onar, S., & Oztaysi, B. (2019). Hospital location selection using spherical fuzzy TOPSIS. In: 2019 conference of the international fuzzy systems association and the European society for fuzzy logic and technology. *Conference: Proceedings of the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019).* Atlantis Studies in Uncertainty Modelling.

Kanj, H., Kotb, Y., Alakkoumi, M., & Kanj, S. (2024). Dynamic decision making process for dangerous good transportation using a combination of TOPSIS and AHP methods with fuzzy sets. *IEEE Access, 12*, 40450-40479.

Kaya, I., & Cinar, D. (2007). Facility location selection using a fuzzy outranking method. *Journal of Mult.-Valued Logic & Soft Computi, 14*, 251-263.

Klaus, B. (1992). R&D cooperation between firms-a perceived transaction cost perspective. *Management Science, 38*(4), 514-524.

Kutlu Gundogdu, F., & Kahraman, C. (2019). Spherical fuzzy sets and spherical fuzzy TOPSIS method. *Journal of Intelligent & Fuzzy Systems, 36*, 337-352.

Kutlu Gundogdu, F., & Kahraman, C. (2020). Optimal site selection of electric vehicle charging station by using spherical fuzzy TOPSIS method. *Decision Making with Spherical Fuzzy Sets Theory and Applications.* içinde Springer, Cham.

Li, S., & Wang, F. (2020). Research on optimization of improved gray wolf optimization-extreme learning machine algorithm in vehicle route planning. *Discrete Dynamics in Nature and Society*, 1-7.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *31st Conf. Neural Inf. Process. Syst. (NIPS 2017).* Long Beach, CA USA,.

Mahmood, T., Ullah, K., Khan, Q., & Jan, N. (2019). An approach toward decision-making and medical diagnosis problems using the concept of spherical fuzzy sets. *Neural Computing and Applications, 31*(11), 7041– 7053.

Martinez, I., Viles, E., & Olaizola, I. G. (2021). Data science methodologies: Current challenges and future approaches. *Big Data Research, 24*, 1-18.

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

Muñoz, M., Miranda, E., & Sánchez, P. (2017). A fuzzy system for estimating premium cost of option exchange using mamdani inference: Derivatives market of Mexico. *International Journal of Computational Intelligence Systems, 10*(1), 153.

Nagpal, G., Ray, A. K., Kharkwal, N., Jasti, N. V., & Nagpal, A. (2023). Challenges in adoption of business analytics by small retailers: An empirical study in the Indian context. *International Journal of E-Adoption, 15*(2), 1-14.

Oguztimur, S. (2011). Why fuzzy analytic hierarchy process approach for transport problems? *ERSA conference* (s. 438). European Regional Science Association.

Oliveira, L., Morganti, E., Dablanc, L., & Oliveira, R. (2017). Analysis of the potential demand of automated delivery stations for e-commerce deliveries in belo horizonte, brazil. *Research in Transportation Economics*.

Onat, A., & Kactioglu, S. (2020). Supplier selection by fuzzy AHP And fuzzy TOPSIS method: An application in retail sector. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 19*(37), 65-79.

Ozdemir, C., Cevik Onar, S. .., & Ekmekcioğlu, Ö. (2024). Innovative order delivery success prediction in online retail: Integrating ML and LLM to gain actionable and understandable insights. *Intelligent and Fuzzy Systems. INFUS 2024. Lecture Notes in Networks and Systems,.* içinde Springer, Cham.

Ozdemir, C., Cevik Onar, S., & Bagriyanik, S. (2020). Estimating shopping center visitor numbers based on various environmental indicators. In C. C.

Kahraman, *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions* (Vol. 42, pp. 1-14). Springer.

Ozdemir, C., Onar, S. C., Bagriyanik, S., Kahraman, C., Akalin, B. Z., & Öztayşi, B. (2021). Estimating shopping center visitor numbers based on a new hybrid fuzzy prediction method. *Journal of Intelligent & Fuzzy Systems*, 1-14.

Ozemre, M., & Kabadurmus, O. (2020). A big data analytics based methodology for strategic decision making. *Journal of Enterprise Information Management*, *33*, 1467–1490.

Oztaysi, B., Cevik Onar, S., & Kahraman, C. (2018). Fuzzy multicriteria prioritization of Urban transformation projects for Istanbul. *Journal of Intelligent & Fuzzy Systems , 30*(4), 2459-2474.

Peng, X., Li, X., & Yang, X. (2022). Analysis of circular economy of ecommerce market based on grey model under the background of big data. *Journal of Enterprise Information Management, 35*, 1148–1167.

Raji, M. A., Olodo, H. B., Oke, T. T., Addy, W. A., Onyeka, C. O., & Oyewole, A. T. (2024). Real-time data analytics in retail: A review of USA and global practices. *GSC Advanced Research and Reviews, 18*(3), 59-65.

Rao, S., Rabinovich, E., & Raju, D. (2014). The role of physical distribution services as determinants of product returns in internet retailing. *Journal of Operations Management.*

Rooderkerk, R. P., DeHoratius, N., & Musalem, A. (2022). The past, present, and future of retail analytics: Insights from a survey of academic research and interviews with practitioners. *Production and Operations Management*, *31*(10), 3727-3748.

Sagar, S. (2024). The impact of digital transformation on retail management and consumer behavior. *IOSR Journal of Business and Management, 26*(1), 6-14.

Sarkar, B., & Biswas, A. (2021). Pythagorean fuzzy AHP-TOPSIS integrated approach for transportation management through a new distance measure. *Soft Computing, 25*, 1-17.

Seetharaman, A., Niranjan, I., Tandon, V., & Saravanan, A. S. (2016). Impact of big data on the retail industry. *Corporate Ownership & Control, 14*(1), 506-518.

Shankar, V. (2019). Big data and analytics in retailing. *NIM Marketing Intelligence Review, 11*(1), 36 - 40.

Singh, R. K., Modgil, S., & Shore, A. (2023). Building artificial intelligence enabled resilient supply chain: a multi-method approach. *Journal of Enterprise Information Management*.

Sund, A., Bjerkan, K., & Hjelkrem, O. (2021). E-groceries: Sustainable last mile distribution in city planning. *Research in Transportation Economics, 87*.

Şekerkaya, A., & Cengiz, E. (2010). Kadın tüketicilerin alışveriş merkezi tercihlerinin belirlenmesi ve bir pilot araştırması. . *Öneri , 9*(34), 41-55.

Thomas, A., & Jose, S. (2023). Analysis of transportation problem under spherical fuzzy environment. *AIP Conference Proceedings.*

Wang, L., Fan, H., & Wang, Y. (2018). Sustainability analysis and market demand estimation in the retail industry through a convolutional neural network. . *MDPI Sustainability,* 1-19.

Wesley, S., LeHew, M., & Woodside, A. G. (2006). Consumer decision-making styles and mall shopping behavior: Building theory using exploratory data analysis and the comparative method. . *Journal of Business Research , 59*(5), 535–548.

Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications,.* New York: The Free Press.

Williamson, O. E. (1981). The economics of organization: The transaction cost approach. *American Journal of Sociology, 87*(3), 548-577.

Yager, R. (2013). Pythagorean fuzzy subsets. *Joint IFSAWorld Congress and NAFIPS Annual Meeting (IFSA/NAFIPS.*

Yuan, Y., Li, H., & Ji, L. (2021). Application of deep reinforcement learning algorithm in uncertain logistics transportation scheduling. *Computational Intelligence and Neuroscience*, 1-9.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338–353.

Zolfani, S. H., Aghdaie, M. H., Derakhti, A., Zavadskas, E. K., & Varzandeh, M. H. (2013). Decision-making on business issues with foresight perspective; an application of new hybrid MCDM model in shopping mall locating. *Expert Systems with Applications, 40*(17), 7111–7121 .

CURRICULUM VITAE

Name Surname **:Çağatay Özdemir**

EDUCATION **:**

- **B.Sc.** **:** 2015, Istanbul Technical University, Faculty of Management, Industrial Engineering Department
- **M.Sc.** **:** 2019, Istanbul Technical University, Graduate School, Industrial Engineering Department
- **MBA** **:** 2021, Quantic School of Business and Technology, Graduate School, Business Administration Department

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2025-Present MEXT AI Center Director
- 2021-2025 VNGRS Data & AI Product Lead
- 2018-2021 TURKCELL Assoc. Principal Data Monetization Product Manager
- 2018-2018 YAPI KREDİ Senior Mass Segment Specialist
- 2015-2018 TÜRK TELEKOM Cloud & ICT Product Manager

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Ozdemir, C., Cevik Onar, S., Con, M.** A Hybrid Approach to Delivery Optimisation Using Machine Learning and Spherical Fuzzy TOPSIS, *Journal of Mult.-Valued Logic & Soft Computing (Accepted : December 2024)*
- **Ozdemir, C., Cevik Onar, S., Ekmekcioglu, O.** 2024: Innovative Order Delivery Success Prediction in Online Retail: Integrating ML and LLM to Gain Actionable and Understandable Insights. In: Kahraman, C., Cevik Onar, S., Cebi, S., Oztaysi, B., Tolga, A.C., Ucal Sari, I. (eds) Intelligent and Fuzzy Systems. INFUS 2024. Lecture Notes in Networks and Systems, vol 1089. Springer, Cham.
- **Ozdemir, C., Cevik Onar, S.,** 2023: Multi-criteria Fuzzy Decision-Making Techniques with Transaction Cost Economy Theorem Perspectives in Product Launching Process. In: Calisir, F. (eds) Industrial Engineering in the Age of Business Intelligence. Lecture Notes in Management and Industrial Engineering. Springer, Cham.

- **Ozdemir, C., Cevik Onar, S., Bagriyanik, S., Kahraman,C., Akalin, B., Oztaysi, B.** 2021. Estimating Shopping Center Visitor Numbers Based On A New Hybrid Fuzzy Prediction Method, *Journal of Intelligent & Fuzzy Systems*, 42(7), 1-14.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- **Ozdemir, C., Cevik Onar, S., Bagriyanik, S.** 2020: Estimating Shopping Center Visitor Numbers Based on Various Environmental Indicators. In: Kahraman, C., Cevik Onar, S., Oztaysi, B., Sari, I., Cebi, S., Tolga, A. (eds) Intelligent and Fuzzy Techniques: Smart and Innovative Solutions. INFUS 2020. Advances in Intelligent Systems and Computing, vol 1197. Springer, Cham.

- **Ozdemir, C., Yildirim, S.** 2021. BİR İLAÇ SORGULAMA SİSTEMİ. Patent numarası: 2021-GE-870013