KADİR HAS UNIVERSITY
SCHOOL OF GRADUATE STUDIES
DEPARTMENT OF ADMINISTRATIVE SCIENCES

THE EFFECTS OF DIGITAL STRATEGIES ON CUSTOMER CHURN IN THE TELECOM INDUSTRY

ÖMER BUĞRA KIRGIZ

MASTER THESIS

İSTANBUL, JUNE, 2022

Ömer Buğra Kırgız

Master Thesis

2022

THE EFFECTS OF DIGITAL STRATEGIES ON CUSTOMER CHURN IN THE TELECOM INDUSTRY

ÖMER BUĞRA KIRGIZ
ADVISOR: ASST. PROF. MELTEM KIYGI ÇALLI

A thesis submitted to
the School of Graduate Studies of Kadir Has University
in partial fulfilment of the requirements for the degree of
Master Thesis in Business Intelligence and Analytics

İstanbul, June, 2022

APPROVAL

This thesis titled THE EFFECTS OF DIGITAL STRATEGIES ON CUSTOMER CHURN IN TELECOM INDUSTRY submitted by ÖMER BUĞRA KIRGIZ in partial fulfillment of the requirements for the degree of Master of Arts / Master of Science/ Doctor of Philosophy in (Name of the Program) is approved by

Asst. Prof. Meltem Kıygı Çallı (Advisor) ....
Kadir Has University

Asst. Prof. Petek Tosun ....
Kadir Has University

Dr. Abdullah Önden ....
Yalova University

I confirm that the signatures above belong to the aforementioned faculty members.

Prof. Dr. Mehmet Timur Aydemir
Director of the School of Graduate Studies
Date of Approval: 23/06/2022

DECLARATION ON RESEARCH ETHICS AND PUBLISHING METHODS

I, ÖMER BUĞRA KIRGIZ; hereby declare

- that this Master of Science Thesis that I have submitted is entirely my own work and I have cited and referenced all material and results that are not my own in accordance with the rules;
- that this Master of Science Thesis does not contain any material from any research submitted or accepted to obtain a degree or diploma at another educational institution;
- and that I commit and undertake to follow the "Kadir Has University Academic Codes of Conduct" prepared in accordance with the "Higher Education Council Codes of Conduct".

In addition, I acknowledge that any claim of irregularity that may arise in relation to this work will result in a disciplinary action in accordance with the university legislation.

Ömer Buğra Kırgız

____

23/06/22

ACKNOWLEDGEMENT

This research study is made possible thanks to several people and institutions. I want to first thank Dr. Öğr. Üyesi Meltem Kıygı Çallı, who is my dissertation supervisor in this research, for her always-on collaboration, openness, and patience in this study. Her support saved this study from getting stuck at certain steps along the way.

I want to thank also Dr. Öğr. Üyesi Sendi Çağlıyor for her expertise in the machine learning field. Her contribution to the analysis part added great value to this study.

I want to thank Turkcell Teknoloji A.Ş. and especially İdris Bozkurt, Koray Çetingöz, and Müşerref Eda Erdemi. They played a vital role in making this study happen by their collaboration on this research study.

Last but most important, I want to thank my wife Büşra Kırgız who supported me right from the outset until the end of this study.

THE EFFECTS OF DIGITAL STRATEGIES ON CUSTOMER CHURN IN THE TELECOM INDUSTRY

ABSTRACT

The telecom industry has been saturated over the last years and organic growth in the number of customers has been slowing down. Institutions allocate a significant amount of resources to reducing churn rates as the variations in service offerings become subtle. Customer retention strategies such as customer relationship management, loyalty programs, and convergence of services are some of the widely-used efforts in the telecom industry in this respect. Thanks to the increasing app penetration, digital loyalty apps, and over-the-top media services emerged as a way of both service differentiation points as well as customer retention strategies. Regardless of all these strategies, some customers will still churn; therefore, churn prediction plays an essential role in the sustainable future of businesses. Churn prediction is used both to detect customers with a high propensity to churn and to interpret the reasons behind the churn decision of customers. This study examines the variables playing important role in churn decisions and the effectiveness of digital loyalty and over-the-top service strategies on customer retention in light of the relationship marketing strategy. The customer churn data in this study is received from a telecom company and contains the attributes of both churner and non-churner customers. Random Forest and Logistic Regression classifiers are used as the machine learning algorithm in the churn prediction model. To understand the variable importance, mean decrease in impurity and mean decrease in model accuracy using permutation are used. The key findings of this research revealed that while digital loyalty app strategies are effective, over-the-top media service strategies play an unimportant role in the churn decision of customers.

Keywords: churn prediction, machine learning, telecom, digital strategy, loyalty, customer retention

TELEKOM SEKTÖRÜNDE DİJİTAL STRATEJİLERİN
MÜŞTERİ KAYBINA ETKSİ

ÖZET

Telekom sektörü son yıllarda doygunluğa ulaşmış durumda ve müşteri adetlerindeki organik büyüme yavaşlama gösteriyor. Hizmet tekliflerindeki farklılıkların da belirsizleşmesiyle birlikte, kurumlar önemli ölçüde kaynaklarını müşteri kaybını önlemeye ayırıyor. Müşteri ilişkileri yönetimi, sadakat programları ve servis birleşmeleri gibi müşteri tutundurma stratejileri, telekom endüstrisinde bu konuyla ilgili uygulanan yaygın eforlar arasında yer almaktadır. Artan telefon uygulamaları kullanımı sayesinde, dijital sadakat uygulamaları ve literatürde over-the-top olarak geçen medya servisleri, servis farklılaşma noktaları ve müşteri tutundurma stratejileri olarak ortaya çıktı. Tüm bu stratejilere rağmen, bazı müşteriler hala telekom şirketlerini terk ediyorlar. Bu yüzden müşteri kaybı tahminlemesi, şirketlerin sürdürülebilir gelecekleri için elzem bir rol oynamaktadır. Müşteri kaybı tahminlemesi, hem kaybedilme ihtimali yüksek olan müşterileri saptamada hem de kaybın nedenlerini anlamada kullanılmaktadır. Bu araştırma, müşteri kaybında önemli rol oynayan değişkenleri ve dijital uygulama ve over-the-top medya servis stratejilerinin müşteri tutundurmadaki etkisini, ilişkisel pazarlama stratejisi ışığında incelemektedir. Bu araştırmadaki müşteri kaybı datası bir telekom şirketinden alınmıştır. Data, kaybedilen ve kaybedilmeyen müşterilerin çeşitli değişkenlerini ihtiva etmektedir. Müşteri kaybı tahmin modelinde, Rastgele Orman ve Lojistik Regresyon makine öğrenmesi algoritmaları kullanılmıştır. Değişkenlerin önemini anlamak için, homojenlikteki ortalama düşüş ve model doğruluğundaki permutasyonlu ortalama düşüş yöntemleri kullanılmıştır. Bu çalışmadaki sonuçlar gösterdi ki, müşteri kaybını önlemede, dijital sadakat uygulamalar stratejisi önemli bir etken iken, over-the-top medya servisler stratejisi bir etki yaratmamaktadır.

Anahtar Sözcükler: müşteri kaybı tahminleme, makine öğrenmesi, telekom, dijital strateji, sadakat, müşteri tutundurma

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The telecom industry has long been known for its monopolistic ecosystem. Early innovators and investors benefited greatly from being the first in the market as well as their infrastructure investments creating high barriers to prospective market entrants. However, the industry has undergone a series of long transformations since the early 2000s paving the way for a more competitive landscape.

According to the 2021Q2 report from BTK (Bilgi Teknolojileri Kurumu), a regulatory institution in Turkey, the number of mobile subscribers is 84.6 million, with a 101.2% penetration. If we were to exclude M2M (machine-to-machine) subscribers and the population in the 0-9 age group, the penetration would result in as high as 109.4%. Although the numbers are likely to vary slightly across different countries, these numbers actively demonstrate that the telecom market is well saturated. The days where organic growth from increasing penetration is long gone and the operators are forced to play a fierce game to capture and retain the subscribers. To put it another way, Kim and Yoon (2004) claims that focus of telecom companies changed from acquiring new customers to maintaining an existing customer base, which is expected in the subscription-based business models. The more the competition increases, the more commoditized the industry’s main services become, which eventually took place in the telecom. Therefore, churn management has become a vital need for a sustainable future for telecom companies (Zhao et al. 2021).

Nowadays, the churn rate among mobile telecommunication companies ranges from 20% to 40% around the globe (Ahn et al. 2006). The milestone might be the introduction of mobile number portability (MNP), which means switching to another telecom operator with your mobile number. MNP was a revolutionary step toward making the telecom industry more competitive. Customers used to stick to their operators even though they were unsatisfied with the service. MNP helped establish a more competitive environment, thus benefiting customers. Attenborough et al. (1998) come up with an estimation of the incremental benefit of introducing MNP in Hong Kong. The introduction of MNP took place in the early 2000s in many countries. Turkey was not among the earliest to implement MNP. It came into effect as late as 2008 (Durukan et al. 2011).

Churn is the aggregate result of a long relationship between a subscriber and a telecom company. The telco deploys specific strategies to increase loyalty and decrease propensity to churn. These could be labeled as precautionary, occurring before any churn probability. There are also retention strategies that are directed at those who are identified as potential churners. Of the many aspects for successful churn management, churn prediction is particularly significant since the potential churners are at the final stage of their lifecycle, and it will be much costlier to win them back if they happen to change the operator. Companies allocate significant resources to churn management to minimize their churn rates.

Moreover, Ganuza and Viecens (2014) claim that new products and services that have been emerging due to the development and increased penetration of the internet had created a competition challenge facing traditional telecom companies. These new players are called OTT, which stands for over the top. Although telecom companies benefited partially from increased data consumption, these new players pose severe long-term risks. Some OTT products constitute a direct substitute to the core products of telecom, like Skype or Whatsapp; some other products mainly focusing on content like Netflix or Youtube. Telecom companies started to different strategies to combat this challenge. Creating their own OTT products, partnering directly with big OTTs, or bundling them in tariff plans are some of the examples emerging in the industry. How OTT strategies intersect with churn management for telecom companies remains an important question yet to be discovered.

2. LITERATURE REVIEW

The purpose of this chapter is to provide the literature review relevant to this thesis. It begins with relationship marketing, customer relationship management, and loyalty. Next is churn definition and churn management. Lastly, over-the-top (OTT) services are discussed. After which, the chapter ends with the research questions and hypotheses.

2.1 Relationship Marketing

Relationship marketing (RM) is a broad sub-field of marketing that refers to the efforts to establish, maintain and enhance relationships with customers and other parties involved so that the objectives of the parties involved are met (Grönroos 1994). Schneider (1980) states that before the 1980s, the number one priority for most of businesses was acquiring customers and the academic literature was shaped accordingly. Retaining the existing customers and strengthening bonds with the company were regarded as idle strategies and often not given much attention as a result.

The term itself -relationship marketing- made its first appearance in the academic literature as early as 1983 by L.L. Berry at an American Marketing Association services marketing conference. Although the term was fairly new at the time, the concept and its practices were utilized by businesses and merchants for a long time.

Das, K. (2009) examines the 209 papers released in years from 1994 to 2006 that have connections with the topic of RM and suggested a categorical classification of the available literature at that time, -objectives, defining constructs, instruments, industrial applications, and issues. He also analyzes the distribution of the years in which the papers were published (Das 2009). The grouping of the years made in this research is 1994-1997, 1998-2001, and 2002-2006. As a result, 62.68% of the papers were published from 2002 to 2006, which implies that the academic research on RM had accelerated in the most recent period.

As it can be seen from the definition of RM, Customer Relationship Management (CRM) and RM are often used interchangeably (Parvatiyar and Sheth 2001), which in result leads to confusion. The confusion stems mainly from the similar nature of these concepts, which involves managing relationships with business stakeholders.

Frow and Payne (2009) suggest a segregation between RM, CRM, and customer management. RM is a more extensive concept that embodies CRM in itself. Therefore, both CRM and RM engage in similar strategies but they are different when it comes to their scopes. CRM is more customer-focused, whereas relationship marketing takes into consideration various stakeholders of companies, not limited to customers. In summary, it could be claimed that CRM is one of the instruments of RM. To put it differently, while RM is comparably more strategic, CRM efforts are considered relatively more tactical (Ryals and Payne 2001).

2.2 Customer Relationship Management (CRM)

Customer Relationship Management began to appear more recently, which is in the 1990s, than RM, which is in the 1980s. As opposed to RM, although there is no concrete trace for the first official usage of the term CRM, it was first seen to be used in IT communities and in the context of technology services (Frow and Payne 2009).

Both the academic literature and the business world have failed to unite around an encompassing definition that would shed light on the implications of and expectations from CRM. Therefore, the term has come to mean many things to many people (Grabner-Kraeuter and Moedritscher 2002).

The relationship between CRM and data are referred in many definitions of CRM. Ryals and Payne (2001, 3) suggest a definition as "information-enabled relationship marketing." Peppers and Rogers (1993) suggest a definition of CRM that could be summarized as one-to-one relationships with customers by processing customer data technologies for long-term sustainable customer retention and therefore business success.

However, narrowing down CRM scope solely to technology solutions as well as tactical actions is a typical mistake that is responsible for the failure of CRM strategies most of the time. Selland and Pockard (2003) claim that the most remarkable outcome could only be expected from CRM on the condition that the organization's orientation is rather a strategic one and this requires a consideration of putting strategy before software.

Hence, before outlining and defining what CRM is and listing its aspects for organizations, it is noteworthy to mention its close relationship with technology. This close relation is regarded as one of the critical factors causing the failure of CRM because it is responsible for the misinterpretation by organizations. In a study by Frow and Payne (2009, 7), the interview with the executives about CRM shows some scattered approaches to the topic:

> To some, it meant direct mail, a loyalty card scheme, or a database, whereas others envisioned it as a help desk or a call center. Some said that it was about populating a data warehouse or undertaking data mining; others considered CRM an e-commerce solution, such as the use of a personalization engine on the Internet or a relational database for SFA (sales force automation. (Frow and Payne 2009, 7)

This actively demonstrates that in the absence of a proper and widely agreed-upon clear definition of the term CRM, organizations tend to associate CRM with technology-based marketing tactics. As a result, this leads to the failure of CRM projects. Kale (2003) highlights this close but misleading fact as a key reason for CRM failure. Similarly, Thakur, Summey, and Balasubramanian (2006) suggest that organizations should take a strategic approach to achieve the best outcomes out of CRM projects. Additionally, Boulding et al. (2005) claim that strategy is the essential aspect of a proper CRM implementation.

The fact that defining CRM is an important initial step towards a successful implementation necessitates a thorough definition. The definition put forth by Frow and Payne (2009, 7) is quite embracive and touching on key areas such as strategy, integration across functions within an organization, customer knowledge, etc.:

> CRM is a cross-functional strategic approach concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. It typically involves identifying appropriate business and customer strategies, the acquisition, and diffusion of customer knowledge, deciding appropriate segment granularity, managing the co-

> creation of customer value, developing integrated channel strategies, and the intelligent use of data and technology solutions to create superior customer experiences. (Frow and Payne 2009, 7)

CRM engages more in the retention of acquired customers. Therefore, the concept of churn and loyalty are the two important phenomena, which we will discuss in the following sections.

2.3 Loyalty

The concept of loyalty can be traced back to the consumer behavior theory. It is the ultimate result that RM and CRM strategies are deployed. Loyalty is rather a psychological and therefore emotional value of the brands is created in customers' minds and hearts. Oliver (1997, 392) defines customer loyalty as "a deeply held commitment to rebury or re-patronize a preferred product/service consistently in the future, thereby causing same repetitive brand or brand set purchasing, despite situational influences and marketing efforts; having the potential to cause switching behavior."

Uncles, Dowling, and Hammond (2003) suggest that loyalty is what customers feel about and show certain behaviors toward brands, products, and services over some time. Mellens, DeKimpe, and Steenkamp (1996) point out that Jacoby and Chestnut's (1978, 80) definition of brand loyalty as the most comprehensive one. Based on the definition, brand loyalty is: "The biased behavior response, expressed over time, by some decision-making unit concerning one or more alternative brands out of a set of such brands, and is a function of psychological (decision-making, evaluative) processes" (Jacoby and Chestnut 1978, 80). The decision-making unit referred to in the definition of loyalty might be an individual, a corporation, or a social group. Loyalty is mostly measured by the purchase or re-purchase patterns or frequency of the decision-making unit. Loyalty is often linked to a brand in the marketing literature. Hence, it is important to define what a brand means to understand its association with loyalty. In Principles of Marketing (2003) book, Philip Kotler and Gary Armstrong points out brand as a combination of a name, logo, terms, signs and symbols differentiating the manufacturer or seller the product.

Although loyalty is often linked to a brand, there is a subtle conceptual difference between brand loyalty and customer loyalty. Brand loyalty refers to the preference of consumers to purchase products and services from a certain brand over others (Ishak and Ghani 2013). It is how the consumer perceives a brand, its image, and its values altogether. It takes place when the consumer chooses a specific brand because he or she knows that the brand offers the value (s)he seeks, whether it is quality, exclusivity, promotion, etc. Customer loyalty, however, is more about the re-purchase of products as well as the frequency of re-purchase. It could be argued that customer loyalty occurs even in the presence of situational influences as well as the marketing efforts of competitors.

Brand building is at the heart of loyalty management. Brand image is essential for constructing sustainable brand equity, ultimately shaping brand loyalty. Faircloth, Capella, and Alford (2001, 61) define brand equity from a consumer-based point of view as "biased behavior a consumer has for a branded product versus an unbranded equivalent." Brand equity is initially researched as "a financial instrument for capturing and measuring the value of brands" (Faircloth, Capella, and Alford 2001, 61).

However, solely using brand equity as a financial measurement instrument would not contribute much to managerial implications. Keller (1993) suggests that it is vital to assess brand equity as to how consumer behavior is shaped by marketing actions that reinforce brand image and brand awareness. It could be summarized that the incremental favorable behavior of consumers towards a certain brand is brand equity and the term's real value lies in understanding these changes in behavior. Keller (1993) sees the brand image as the consumers' perception due to associations with the brand.

In line with what Keller (1993) states, Aaker (1991) claims that brand equity is the overall result of brand image, brand loyalty, brand awareness, perceived quality, and other brand assets (Faircloth, Capella, and Alford 2001). Therefore, marketers need to enhance brand image and brand loyalty if they seek better brand equity. Faircloth, Capella, and Alford (2001) research about this to shed light on the topic that has been conceptualized but not materialized as much. In the research, the effect of various brand associations on brand image is examined. These associations results in a positive brand image that in turn increases the purchase intentions as well as inclination to pay premium prices, which is what is referred to as -brand equity-.

Moreover, brand loyalty and therefore brand equity offer more than customer acquisition, purchase frequency, or premium pricing. They enable a form of protection of the customer base against the competition (Kotler and Armstrong 2003). Mascarenhas, Kesavan, and Bernacchi (2006) and Datta (2003) claim that brand loyal are also responsible for a significant reduction of marketing costs, in addition to ensuring increased sales.

This thesis focuses more on loyalty since this study aims to understand telecom customers' churn behavior. Hence, it could be concluded that brand and customer loyalty are the building blocks of strategic marketing and, therefore should be considered as the primary objective. They play a significant role in creating a long-term sustainable competitive advantage, which is essential for sustainable business growth.

2.3.1 Loyalty and churn in the telecom industry

Churn in the telecom industry is a major challenge facing marketers due to intense competition. The industry has long been in a saturation zone, and players actively seek differentiation tactics such as competitive offers to acquire newcomers (Óskarsdóttir et al. 2016).

Keaveney (1995) points out that customer churn erodes profits in telecom companies. New customer acquisition can be more expensive than retention of the existing customer (Athanassopoulos 2000). Hence, churn prediction management gained a great deal of popularity for both researchers both in academia and among companies.

Churn management starts with detecting those customers that are likely to churn. However, churn prediction is a necessary but insufficient step towards holistic churn management, which should be supplemented with effective loyalty-building strategies. Therefore, service providers have increasingly focused on long-term relationships with existing customers in recent decades (Amin et al. 2017).

Ganesh, Arnold, and Reynolds (2000) show that customer loyalty is critical to a sustainable future for organizations since loyal customers are more profitable. The loyal customers buy more at higher prices with little extra investment being required. Moreover, loyal customers create a positive word of mouth effect, bringing future customers.

Additionally, Pareto Principle (also known as the 80/20 rule), suggests 20% of the causes are responsible for 80% of the effects. Applying this rule to churn management, it is necessary to segment users based on the value they add to the organization. In fact, Reichheld and Sasser (1990) state that companies should not try to keep all the customers. If some of them eventually want to leave and they are not profitable, it is best to use resources to serve other profitable customers. Reichheld and Sasser (1990) also state that a customer yields fewer profits in the first years after acquisition but gradually does much more, meaning the longer relations with customers are more profitable for the company. His research suggests that an organization could enjoy anywhere between 25% to 85% more profit if it manages to reduce the churn rate by only 5%. Moreover, every churned customer means missed opportunities for various upselling and cross-selling. This in turn leads to decreased revenue & profits (Zhao et al. 2021).

In summary, it could be argued that there are two main objectives to applying for loyalty programs in the telecom industry. The first and the most obvious one is to drive sales and therefore revenue. Loyalty programs help increase basket sizes, frequencies as well as cross-product opportunities. The second and more related to this research is to defend the customer base against the competition. Creating customer loyalty is considered to enhance the bond between the customer and the brand, therefore the customer is less likely to switch the brand, which is called churn. Both two objectives lead to increased profitability in return.

2.3.2 Types of loyalty strategies

As loyalty is the ultimate goal, it can be argued that all the marketing strategies and activities that are targeted at the existing customer base can be called loyalty strategies. In the history of the telecom industry, various ways have been introduced to increase customer loyalty. Some of the most predominant ones involve customer clubs and bundling strategies.

2.3.2.1 Loyalty programs and customer clubs

Main aim of customer clubs is customer segmentation. Customers are segmented based on certain similar characteristics or attributes and then they become eligible to be given certain exclusive benefits. Most of the time, it’s either age or invoice amount being used to create the customer clubs.

Gustafsson, Roos, and Edvardsson (2004) try to understand the effects of customer clubs on customer loyalty. Following the rise of various telecom operators in Europe, the companies’ objective has not only been to attract new customers but also to retain the existing customers since attracting a new customer is much costlier than retaining one. Also, loyal customers who stay longer are likely to yield more profit in the long run than those temporary ones who are disloyal and inclined to switch to competitors more easily. They suggest that the main idea of customer clubs is to lock the customers into a relationship with the company and to build a switching barrier, since being a member means that future benefits accumulate in the form of rebates, products, or special offerings. Yet there has emerged a challenge for companies as a result of the proliferation of these marketing activities. Customer clubs have become so uncommon that customers started to see these programs as a prerequisite, resulting in efforts being less loyalty building.

Gupta and Sahu (2015) examine the effects of relationship marketing, not only from the perspective of customer clubs but in a broader way of relationship marketing activities on customer loyalty in the telecom industry in India. The research utilized a survey of sales professionals who use mobile services for business usage. It is highlighted in this article that the cost of serving a loyal customer is much less than attracting and serving a new customer, parallel with findings from other studies discussed above. The study is centered around the idea that there is a strong relationship between business performance and relationship marketing in the telecom industry. It’s also mentioned that this relationship is more important and stronger in the telecom industry than in many other industries.

Verhoef (2003) studies the effects of economic incentives as a loyalty and retention strategy. The study found out that there is a positive relationship between the loyalty programs that include economic incentives and these two metrics: customer retention and customer share. The term customer share is defined as: "the ratio of a customer's purchases of a particular category of products or services from supplier X to the customer's total purchases of that category of products or services from all suppliers (Peppers and Rogers 1999 in Verhoef 2003, 30)."

2.3.2.2 Switching barriers

Many service providers rely on switching costs to lock the door for their existing customers. High switching costs create a barrier thus helping increase loyalty even though it is at the expense of customer satisfaction. Lee, Lee, and Feick (2011) study French telecom customers to find out the link between switching costs and customer loyalty. His study reveals that in the case of customer satisfaction remaining stable, higher switching costs increase loyalty and retention, thus decreasing churn. It does not imply any positive correlation between switching costs and customer satisfaction. On the contrary, customers prefer shorter periods of contracts if they are offered the same price. The link between switching costs and loyalty stems from price advantages. Confraria, Ribeiro, and Vasconcelos (2017) find out that customers are willing to pay 1.3 euros monthly to shorten the contract period by 6 months. In the Turkish telecom industry, BTK (2021), which is a government institution regulating the telecom industry, rules out a switching cost algorithm to protect the rights of customers. According to the algorithm, 2 ways of calculation are made to determine whichever method yields less switching cost for the customer and then the one with the less switching cost is applied to the customer. The first method is the sum of discount benefits the customer has received until the time of calculation thanks to the contract he made with the telecom operator. The second method is the sum of the remaining payments owed to the operator based on the contract. For example, if a customer pays 30$ per month for his tariff based on an annual contract, where he would otherwise have paid 50$, and he is in the 5th month of the contract, the switching cost occurs as follows: (50-30)*5 =100 (the first method), 30*7 = 210 (the second method). In this case, the switching cost for the customer is 100$ because it is the lower one of the two methods. However, telecom companies started to increase the difference between the price of a contracted tariff line and the price of the same tariff line without a contract to benefit more from this calculation method. Therefore, this also results in the continuing effect of switching costs on loyalty in the telecom industry despite the efforts of BTK to decrease competition barriers. Kim and Yoon (2004) define switching costs more considerably by incorporating cost types such as loss cost, adaptation cost, and move-in cost. Loss cost is associated with loss in social status or performance. Adaptation is about adapting to the new telecom provider that churned customer moves to. Move-in is related to operational costs associated with churn action.

2.3.2.3 Convergence of services

Thanks to the convergence of services in the telecom industry, product bundling is another loyalty strategy to lock customers in. The innovation and technology evolution that has long been taking place in the telecom industry has led to increasing IP-based networks. Moreover, the penetration of broadband internet services has grown dramatically to an extent where the industry is now called saturated. Also, new competition has arisen from service providers such as TV streaming services. Lastly, increasing the line of service required little infrastructure investment for the telecom companies since they already invest heavily in it. All of these factors drove the industry towards convergence of service. This is also called product bundling. Product bundling is selling several different products as a single unit at a discount price that is less than what would otherwise cost. Thanks to the development of triple play (also known as 3P), bundling strategies gained importance to lock customers in and increase loyalty. Triple-play services mean fixed voice, broadband, and TV being bundled and sold as a single offer. Quadruple play services refer to mobile data or voice in addition to the triple play products. In OECD countries, this convergence of products resulted in increased competition between service providers.

Ferguson and Brohaugh (2008) explain the rising competition in the telecom industry. The need for retaining existing customers is highlighted as it will be more and more costly to acquire new customers in this highly competitive market. The emergence of services like IPTV and VOIP paved the way for new areas for growth for telco companies and the research suggests that companies should use this type of service as a bundling strategy to increase customer satisfaction and loyalty in turn.

2.4 Churn Definition and Its Types

The term churn has a different usage in the field of marketing and business than in daily life. Zhao et al. (2021, 2) put it as “Customer churn refers to the phenomenon that customers no longer buy products or services of enterprises for various reasons.” In the marketing field, churn simply means a loss of customers. However, there are varying definitions of churn not only in the literature but also among telecom organizations as well. Some of the ways in which a customer action might be labeled as churn include MNP (also known as port-out or MNP- mobile number portability), voluntary churn, and involuntary churn. For the telecom industry, the categorization is slightly different from the banking industry in that it has the additional type referred to as MNP. It is because of the gsm number portability option, which the banking industry lacks.

2.4.1 Mnp (port-out)

In short, port-out is switching from one telecom operator to another by carrying your number with you. MNP is a revolutionary innovation that came into effect during the early 2000s across the globe and it transformed the whole marketing and loyalty-building efforts. Before MNP, customers had to terminate the old mobile phone number and create another one at the new service provider. It created a huge switching barrier; thus, favoring those organizations that were earlier in the market and managed to acquire a certain level of penetration.

2.4.2 Voluntary churn

Voluntary churn is the transaction of terminating a subscription at the current service provider and closing the mobile phone number for usage.

2.4.3 Involuntary churn

Involuntary churn occurs when a service provider terminates the subscription as a result of failures of a series of attempts to charge in a certain period of time. This transaction happens based on some rule sets and customers are given certain deadlines before the phone number is to be turned off for incoming or outgoing calls or SMS.

Since we are interested in the effects of relationship marketing strategies on customer loyalty and churn in this study, we believe focusing on the losing customers to another telecom operator prevents digressing from the context. Therefore, this study considers MNP & voluntary churn as the subject of interest regarding to what constitutes churn.

2.5 Churn Management and Customer Churn Prediction (CCP)

Churn management could be described as the sum of all actions to retain and defend the existing customer base while also considering profitability and sustainability. Starting with the calculation of churn probability for every customer, it involves deployment of retention as well as win-back campaigns to be targeted at them to prevent churn.

Churn management starts with churn prediction. CCP is in essence a classification problem that organizations in varying industries seek to solve to maintain their existing valuable customers. Some of the prominent industries where CCP is highly valued include but are not limited to banking, insurance, internet, and telecom industries (Oskarsdottir et al. 2016).

Churn prediction in the telecom industry attracted more and more attention in the literature (Amin et al. 2019; Luo et al. 2007; Verbeke et al. 2012) mainly for two reasons. Firstly, the growth and the penetration slowed down in the telecom industry, yet competition has become ever fiercer. Moreover, it is now easier than ever to change service providers and the difference in the quality of services' network gradually disappeared in the last decade. This makes telecom products a commodity for consumers. The effect of CCP efforts on organizations' revenue is demonstrated in the literature (Amin et al. 2017; Oskarsdottir et al. 2016). Secondly, increasing data touch points resulted in higher data volume, velocity, and variety available to organizations, which in turn caused the proliferation and popularization of big data management techniques.

The main goal of CCP is to score existing customers based on their propensity levels to churn so that retention and win-back campaigns can be optimized towards a more efficient resource allocation and management. As a result, customers with a high probability of churning would be targeted more intensely, whereas those with less probability of churning would not be bothered or even triggered by win-back strategies. Moreover, campaigns that offer additional discounts might lead to revenue cannibalization if the target audience has no intention of churning in the foreseeable future. Hence, churn prediction techniques should also take profitability into account. Mozer et al. (2000) mention an optimum threshold of churn probability. According to the research, only the customers with higher probability scores than the threshold should be included in retention and win-back campaigns to minimize costs and maximize returns for organizations.

Verbeke et al. (2012) build on top of this and argues that traditional CCP model selection procedures follow a set of solely statistics-based criteria. This in turn results in suboptimal models when it comes to cost-effectiveness, which is one of the primary objectives that necessitates churn management in the first place. He developed a profit-driven measurement approach by incorporating the calculation of profit of a retention campaign. It is then used to determine an optimal fraction of the scored customers to be included in retention campaigns. The study concluded that the new measurement method generates a higher profit than standard statistical methods.

The literature about customer churn prediction in the telecom domain could be categorized into two as: the determinants of churn (or churn reasons) and model building and selection.

2.5.1 Features of customer churn prediction (determinants of churn)

Determining the features responsible for churn behaviors carries a great deal of value as it would lead to building a better churn model and taking necessary cautionary steps to prevent churn in the future. Classifiers used in the churn prediction models reveal which variables are of more excellent value than others.

It is important to recognize that even if not all the variables, some of the variables in the prediction model should have a statistically significant effect on churn behavior to predict effectively. If the major variables affecting churn are not present, the accuracy of the model would decrease and it would be harder to take preemptive steps for future churners. That is a challenge for any model building. For example, if a major network issue took place in a certain month and it led to a huge flock of subscribers leaving the company, any model to be built by the data from that month would lead to inferior or modest predictions.

Moreover, Verbeke et al. (2012) claim that business acumen is required to evaluate the overall resonance of the model with the business. Therefore, business experts should be involved in the variable selection processes. He claims that dumping all the variables available at discretion would hinder the overall performance of the model because some of the variables that could be labeled as a dominant important feature might just be a symptom of the churn behavior, rather than a cause.

Kisioglu and Topcu (2011) implement a churn prediction model using the Bayesian Belief Network as the model of interest. The data used in this study is provided by a Turkish telecom service provider, similar to our research. The research shows that minutes of calls, billing amount, the frequency of calls to people from different providers, and the tariff type are the predominant variables affecting churn behaviors.
Contract information is also critical because service providers rely heavily on long-term contracts to lock customers in, thus increasing switching costs, as mentioned in the loyalty strategies section above. Call center logs and complaint information are also noteworthy because these signal a red flag in terms of customer satisfaction.

According to Oskarsdottir et al. (2016), studies show that incorporating social network variables into the models increases the overall performance of churn prediction models. The main goal is to involve the social influence that the churned customers have on others, which could be called as word of mouth effect. By using call detail records, the social network mapping is created and the connection strength between churners and their connections is determined. Similar to social network analysis, the scale of the total network that a telecom provider has also plays an important role when deciding to choose a service provider, according to Kim and Kwon (2003). Their research also suggests that the bigger the scale of a network is, the more favorable to customers it is. The result of the research shows that intra-network call discounts and call quality are partly responsible for this behavior. Confraria, Ribeiro and Vasconcelos (2017) find out that customers are willing to pay 2.5 euros per month to be a part of a larger network.

Switching cost is an important factor affecting churn propensity in an inversely correlated way. As mentioned above previously, Lee, Lee and Feick (2011) show that the higher the switching cost, the higher the customer loyalty and the lower the churn propensity, assuming constant customer satisfaction. In addition to hard switching costs which are reflected as a cash penalty to be paid by customers, telecom companies also engage in loyalty programs with economic incentives to intensify the switching effect on customers. Giving away free GPRS, minutes for the call, and membership programs for exclusive places are some of the most prevalent examples.

Studies have shown that customer satisfaction has a significant impact on customer churn (Uner, Guven, and Cavusgil 2020; Ahn et al. 2006; Gronholdt, Martensen, & Kristensen 2000; Kim & Yoon 2004). Uner et al. (2020) and Gerpott et al. (2011) indicate that an increase in customer satisfaction results in a significant contribution to customer loyalty; thus, enhancing lowering customer churn.

Expenditures a customer makes at his telecom service provider are identified as an impactful parameter in the literature. This expenditure metric is also known as revenue per user or revenue per subscriber. Madden, Savage, and Coble-Neal (1999) find out that ISP (internet service provider) expenditures have a positive impact on churn behavior. However, household income is found inversely related in this research, which means the more the expenditure and the less the household income available, the more likely to churn a customer is.

Zhao et al. (2021) contextualize the literature on churn reason in further detail and then comes up with a categorization while trying to build the research hypothesis. He categorizes scholars' literature based on these 3 aspects; consumption-related variables, customer socio-economic variables, and lastly enterprise-related variables. The first two variables are more obvious in that the first category includes call duration and consumption amounts and the second includes age, gender, household income, etc. Some of the variables that fall under the last category are enterprise channel operation ability and purchase of certain products.

2.5.2 Churn prediction models (model selection)

Churn prediction is essentially building a model trained to label the current customer base as prospective churners and non-churners. Verbeke et al. (2012) make a thorough categorization of churn classifications (i.e., models) available in literature.

Choosing the best CCP model for churn prediction is an optimization problem and comparative studies have been made to address this. Huang, Kechadi, and Buckley (2012) compare 7 prediction models (Logistic Regression, Linear Classifications, Naive Bayes, Decision Trees, Multilayer Perceptron Neural Networks, Support Vector Machines, and the Evolutionary Data Mining Algorithm) by looking at the AUC scores and suggest that the selection of best prediction model depends on the objective of the decision-makers.

However, Verbeke et al. (2012) claim that the results of these comparative studies are often conflicting with each other. It is referred to the studies of Mozer et al. (2000) and Hwang et al. (2004) in which the former finds out neural networks outperforms logistic regression when it comes to prediction of customer churn and the latter states the opposite. According to Verbeke et al. (2012), this may stem from the fact that comparative studies are often run by a limited number of classifications on a single data set. The nature of data has potential to affect the comparison metrics.

The success of CCP models is mostly measured by certain statistical metrics such as the AUC score. However, at the heart of CCP approaches lies profitability. CCP models are built to prevent churn because the cost of acquiring a new customer outweighs the cost of serving the existing customers. Therefore, Verbeke et al. (2012) develop a profit-centric performance measure where the model is optimized with the goal of maximum profit and not every customer whose likelihood of churning is high is eligible to be a target audience of retention campaigns. The results indicate a strong impact of the new profit-centric measure on the profit yielded from a retention campaign. In this research, Verbeke et al. (2012) find out that the Decision Trees classifier (i.e., model) provide the best performance despite a not-substantial difference, which makes Verbeke et al. (2012, 227) state that "properties of modeling techniques besides the predictive power have to be taken into account when choosing a classification technique, such as comprehensibility and operational efficiency."

It could be argued that CCPs are built with mainly two objectives. The first is to detect customers that are more likely to churn. However, a CCP model without revealing any insight as to reasons leading to churn behavior is of little use in marketing. Therefore, the second purpose is to understand which features have the most important in detecting churners so that following marketing strategies and actions can be formed with the outputs.

Saarela and Jauhiainen (2021) suggest that there is usually a trade-off between the two objectives. According to Saarela and Jauhiainen (2021), simple linear models are effective in understanding and interpreting the churn reasons; however, they perform poorer as compared with non-linear ones. More complex models are harder to comprehend, yet more effective in achieving higher prediction accuracy. Verbeke et al. (2012) suggest similarly and states that rule induction techniques, decision tree approaches, and classical statistical techniques such as logistic regression and Naive Bayes or Bayesian Networks are more suitable if the primary aim is to understand what drives churn behavior and yet they are still effective in terms of prediction accuracy. Amin et al. (2017) refer to the study of Richter & Slonim (2010) and state that the random forests are criticized for their complexity to interpret and made an inference about the causes of churn.

Kişioğlu and Topcu (2011) state that studies in the literature about CCP in the telecom industry rely on such models as neural networks, decision trees, Naive-Bayes, and cluster analysis techniques but there is no previous study where Bayesian Belief Network is used to predict telecom customer churn. Therefore, Kisioglu and Topcu (2011) build a CCP model based on the Bayesian Belief Network.

While applying binomial logistic regression to his study, Uner, Guven, and Cavusgil (2020) refer to the study of Howitt and Cramer (2011) and state that binomial logistic regression is preferred when the dependent variable is binary because a normal distribution is not needed in this classification method. A binary variable is one with only two values such as yes or no, 1 or 0. This is also called a discrete choice theory. The dependent variable in customer churn prediction is whether a customer churns or not.

Neslin et al. (2006) design a churn prediction tournament where participants try to predict customer churns on publicly available data. Participants are later surveyed about the predictive model as well as feature selection methodologies they use. According to the survey, logistic regression and decision trees are the most used models by participants with 45% and 23% respectively.

Zhao et al. (2021) also use logistic regression as the CCP model to predict customer churn on Chinese telecom data.

2.6 OTT Media Services as New Competitors

OTT services offer content to viewers over the Internet. The type of content could be video, audio, text, or any other communication tool. They do not need to have an infrastructure investment to make streaming possible. Some of the well-known examples of OTT media services are Netflix, Youtube, Amazon Prime TV, Whatsapp, Instagram, and Zoom.

OTT stands for over-the-top, simply as “services carried over the networks, delivering value to customers, but without any carrier service provider being involved in planning, selling, provisioning, or servicing them – and of course without any traditional telco booking revenue directly from them (Green and Lancaster 2006, 2).” They do not pay any platform or regulatory fee for the usage of network infrastructure.

There are two important implications when it comes to the effects of OTT media services on telecom companies. Firstly, OTTs do not create any direct revenue for the telecom companies although they deliver their services through these network companies. However, they are responsible for increases in data consumption since content streaming usually takes heavy data usage. As a result, revenue from data usage increases for telecom companies.

Secondly, the product and offerings of OTT media services are close substitutes to that of telecom companies, thus creating competition for the telco (Sujata et al. 2015). It in turn poses a serious risk for long-term revenue sustainability. The most obvious example of this risk took place for SMS and voice revenue for the telco. The two important revenue sources were cut out due to increased penetration of OTT messaging and voice delivering media services such as Whatsapp. Ganuza and Viecens (2014) and Shin (2012) mention that mobile operators in Korea faced a reduction in their voice revenue as a result of Internet-based calls or mobile applications that provide voice and messaging services over smartphones. Nikou, Bouwman, and De Reuver (2012) state that operators all across Europe faced decreases in their revenue per user ratio due to such OTT communication services as mobile VoIP, Whatsapp, and social media like Facebook and Twitter.

Sujata et al. (2015) study the impact of OTT services on telecom and states that the change facing this industry is led by changing customer preferences. Farooq and Raju (2019) state that telecom companies may benefit from this change if they manage to comprehend and take on customer needs and requirements and produce or sell value-added products or services on top of their core legacy telecom products. This is also known as the bundling strategy. As Sujata et al. (2015) recommend, bundling with OTT services is a well-known strategy that many operators resort to. Some examples mentioned in this study are TeliaSonera bundling Skype, Vodafone UK bundling Spotify, Sky Sports, or Netflix in their data plans and tariffs.

Sujata et al. (2015) also recommend telecom companies partner with OTT services and/or even create their own OTT services. Ganuza and Viecens (2014) discuss the conceptual framework as to how traditional telecom companies respond to the evolution of OTT services in the context of the content that they deliver. The study lays out two main strategies that have been and are being developed by traditional telecom companies; bundling of services (fixed and mobile telephony, Internet and TV) and development of their own OTT services. Sujata et al. (2015) discusses some examples of the in-house development of OTT services by telecom companies. Bobsled by T-Mobile USA, Tu Me by Telefonica Digital offers voice calls and text messages free of charge, and Libon by Orange.

2.7. Research Questions and Hypotheses

Research questions of the current study are given below.

Research question 1: Do OTT apps have an impact on the churn decision of customers?
Research question 2: Does the usage of main digital channels and loyalty apps have an impact on the churn decision of customers?
Research question 3: Do loyalty giveaways have an impact on the churn decision of customers?
Research question 4: Which features have the most impact on the churn decision of customers?

Based on the research questions above, the following hypotheses are developed.

Hypothesis 1: Usage of OTT apps have a significant effect on the churn decision of customers.
Hypothesis 2: Usage of digital channels and loyalty apps have a significant effect on the churn decision of customers.
Hypothesis 3: Loyalty giveaways have a significant effect on the churn decision of customers.
Hypothesis 4: Usage of OTT apps and digital channels are the most important features affecting churn decision of customers.

3. METHODOLOGY

3.1 Research Method and Design

Python is used to prepare and build the churn model in this analysis. The python libraries such as pandas, numPy, sci-kit learn have made data manipulation, preparation, and building of machine learning models very easy for data scientists.
Building a churn model requires several preprocessing steps as it is often the case in data mining that the data that is pulled from databases is not ready to be trained for the classifiers. Some of the most widely utilized techniques are also used in the method of this research. They mainly are; handling missing values, the transformation of categorical variables to a numeric format, finding out variables with one unique value (thus zero importance), detecting highly correlated variables, standardization, and finally splitting the data into test and train samples to leave a hold-out sample for model performance measurement.

Although some of these steps in data preprocessing are not required if the chosen classifier does not use the gradient descent technique in the optimization of classifying, classifiers that use gradient descent based or distance-based optimization algorithms are affected in the absence of preprocesses that are mentioned above. While Linear Regression, Logistic Regression, and Neural Networks use gradient descent optimization; KNN, K-means, and SVM classifiers use distance-based optimization. These are the ones that are most affected by unstandardized features. However, because of the tree splitting optimization the Random Forest classifier does not require standardization or null imputing preprocess techniques. Data preprocessing needs to be performed, especially in cases where more than one classifier is used.

In summary, churn prediction methodology involves; getting sampled data from a database, data preprocessing, fitting a classifier algorithm, and then predicting future events (Huang, Kechadi, and Buckley 2012). The steps in data preprocessing will be discussed in the following sub-section.

3.1.1 Preprocessing

Machine learning algorithms produce the best results if they are fed with good quality data sets as inputs. If data has features that were improperly filled in its source or with many missing values or data includes data types that are not compatible with the classifier, the prediction algorithm is likely to produce inferior performances (Kotsiantis, Kanellopoulos, and Pintelas 2006). Therefore, data preprocessing is vital in machine learning.

Huang, Li, and Xie (2015) find out that data preprocessing techniques have a significant impact on the performance of prediction techniques. According to Kotsiantis, Kanellopoulos, and Pintelas (2006), there is no one size fits all solution to preprocess any data sets; therefore, data scientists should be well aware of the data he/she uses. The final output of the steps in data preprocessing is the training set that a machine learning algorithm will use to train itself.

3.1.1.1 Cleaning out irrelevant or not useful variables

There might be certain variables that essentially do not contribute anything to the model. Some examples are variables with one unique value for all the records, variables that describe another variable in the same dataset in certain unit measures such as kilograms, minutes, gigabytes etc. These features are irrelevant and do not contribute or give any insight. Therefore, these features are dropped from the dataset in most cases. Apart from unit descriptor variables, the variables with only 1 unique numeric value for the entire dataset are usually checked by filtering with standard deviation = 0.

3.1.1.2 Transforming the target value

Although classifiers like Random Forest can work with categorical data, all the inputs need to be numeric when using Logistic Regression as the algorithm is not able to handle categorical variables. However, not only the independent variables but also the dependent variable, in other words - the target value-, could be categorical in databases. Even if the target value is a binary variable with only 2 unique values, its values might be Y and N, rather than 1 and 0. Therefore, the transformation from categorical to numeric needs to be performed to work flawlessly with Logistic Regression.

3.1.1.3 Missing values

Data does not come in the perfect format in the real world and missing values are one of the most prevalent forms in which this problem occurs. Missing values are records that have no value available for various reasons. The data source that feeds the database may sometimes fail to capture the data and then feed the database as a result. Moreover, the database might be designed in a way that a missing value means the absence of that attribute for the relevant record, for which a missing value means 0 zero. Replacing missing values with 0 for this type of variable is a quick solution in the case that the performer is aware of the nature of the data he is handling.

Of the methodologies for handling actual missing values, imputing with median is widely used. Imputing missing values means replacing the missing values with the median value that is calculated with the non-missing part of that variable. Pandas library has a function that is referred to as pandas.DataFrame.fillna. It enables users to fill the missing values with the input provided, for which median selection is possible.

The calculating median is not applicable for the categorical variables. Therefore, eliminating the records with missing values on the categorical variable is rather an easy option if the missing value ratio is not significantly high.

3.1.1.4 Transforming categorical variables

Although Random Forest can work with categorical variables, dummy transformation is needed so that logistic regression can be used as a performance comparison algorithm. Pandas library in Python has a practical function, which is pandas.get_dummies, that assigns arbitrary integer values to the unique values in a variable. It does that by creating new columns with each unique value. Categorical variables are handled by this method.

3.1.1.5 Correlated variables

Detecting collinearity is an important step in creating a successful churn prediction model as correlated variables add little incremental accuracy but make it harder to evaluate the variables with more importance in the model. Multicollinearity heatmap is a widely used tool to visualize the correlation between the variables. An example is shown in Figure 3.1. The degree of correlation among features is assessed based on the value of correlation coefficients placed on each cell.

It is important to note that the heatmap uses a linear correlation type of measurement; in other words, no curve type of correlation implication could be made via this method. Between -1 and +1, the higher the positive linear correlation there is, the closer to +1 the correlation coefficient. The same goes for a negative linear correlation but closer to -1 this time.

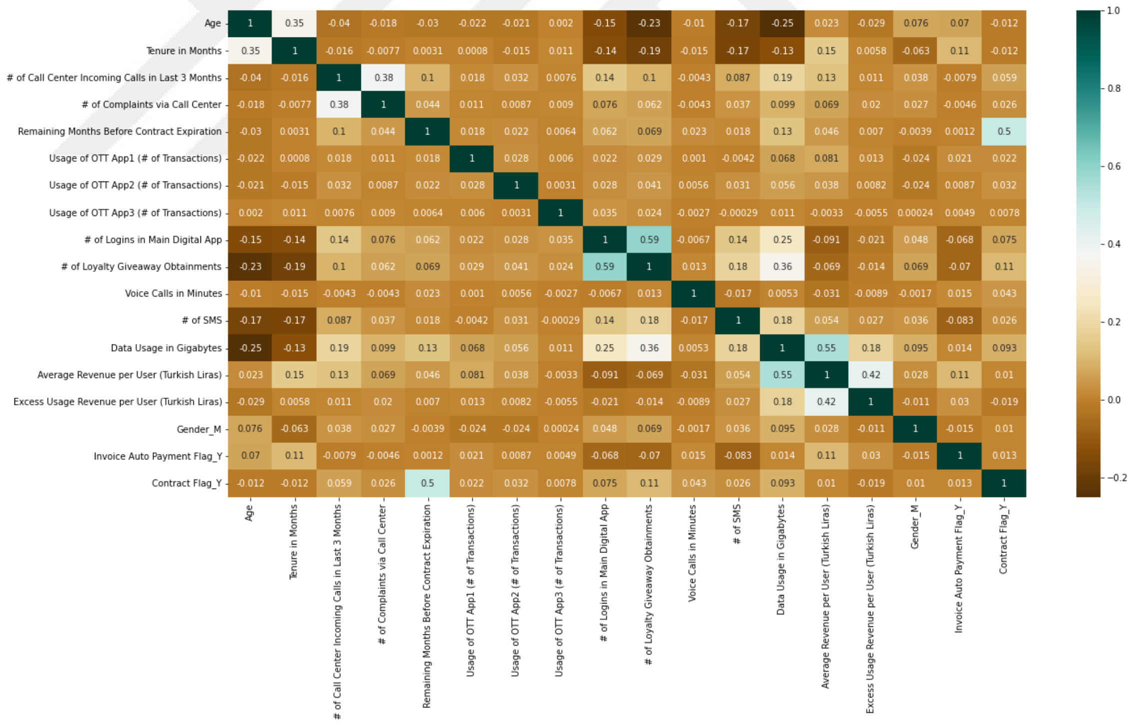

Figure 3.1 Collinearity Heat Map

3.1.1.6 Class imbalance

Class Imbalance is a problem that occurs when the total number of one class in the target is significantly higher than the other in the dataset. This phenomenon creates an optimization problem for the machine learning algorithms because they optimize for maximizing accuracy while minimizing the errors, in other words predicting the churners as true while non-churners as false. A machine learning model that is trained on a dataset with the class imbalance problem would yield very high accuracy in predicting class 0 (say majority class), but perform very poorly predicting and capturing class 1 (say, minority class).

There are ways to overcome the problem of class imbalance. The most common ones are utilizing under-sampling or over-sampling techniques, as shown in Figure 3.2. Chawla et al. (2002) propose an over-sampling approach where synthetic data is generated for the minority class via computing the k-nearest neighbors.

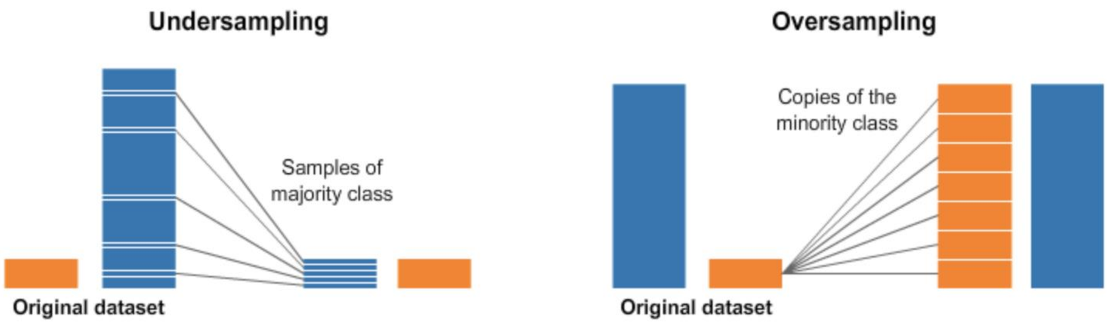

Figure 3.2 Under-sampling and Over-sampling (Liu, Z., Hardy, J., and Zhang, J. 2019)

3.1.1.7 Test and train split

Data needs to be separated into test and training subgroups to prevent overfitting and measure the results in a better way. The machine learning model will learn from the training dataset and then test its performance on the test dataset. The test dataset ratio might change depending on the size of the total data.

3.1.2 Classifier selection

Machine learning algorithms are typically categorized into supervised and unsupervised learning. Supervised learning is an approach where the algorithm learns from training data with labeled input and output and then gets ready to make future output predictions when provided with new inputs. For example, you provide blood pressure and body fat percentage of patients (independent variables) as well as their heart attack history (target value), which makes it labeled data. The algorithm is supervised to learn from this labeled data that high blood pressure leads to a heart attack. Supervised learning algorithms are also further categorized into classification and regression methods. Classifications are generally used when the target value has a few specific categories like fraud/not-fraud, churn/non-churn, apple/orange etc. Support Vector Machines (SVM), Naive Bayes, Decision Tree, and Random Forest algorithms are widely used classification algorithms. Regression-supervised algorithms, on the other hand, are more suitable when the target value is continuous like forecasting sales volume, weather, etc. Some examples are linear and logistic regression as well as polynomial regression. Regression algorithms can also be used to classify target values. Unsupervised learning is another approach in machine learning where data is not labeled and the algorithm finds hidden patterns in the dataset. Unsupervised learning algorithms are further categorized into clustering, association, and dimensionality reduction.

Churn prediction is labeling customers as potential churners and non-churners by a machine learning algorithm that learns from the data of past churners. Past data of those who churned is used to train the model and the model predicts future churners, which makes it a good example for supervised machine learning techniques.
Random Forest is a well-known and accepted supervised machine learning algorithm in churn prediction. The concept is first offered by Tin Kam Ho (1995) using the methodology called the random subspace. This methodology builds on binary decision trees that use a single feature at each nonterminal (decision) node (Ho 1995) by randomly selecting a subset of features, not single. Breiman (2001) later develops an extension where he combines what is referred to as bagging, which is an abbreviation for bootstrap aggregating, and a random selection of features. Subsets of training samples are drawn with replacement (bootstrapping method) to grow each tree where features are also

randomly selected. Samples are generally divided into in-bag samples, which make up two-thirds of the original sample, and out-of-bag samples, which make up one-third. While in-bag samples are used to train trees, out-of-bag samples are used to cross-validate the performance of the Random Forest model (Breiman 2001). Aggregating part occurs when averaging the final vote of all trees for classifying the label. The class that gets the most credit is chosen as the outcome of the Random Forest model.

Sperandei (2014, 12) simply puts Logistic Regression as "Logistic regression works very similar to linear regression, but with a binomial response variable." Logistic regression provides the probability of an outcome (target value) based on independent variables. The algorithm can be written as:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_m x_m \tag{3.1}$$

"where $\log\left(\frac{\pi}{1-\pi}\right)$ is the probability of an event, and β_0 is the regression coefficients associated with the reference group and the x_1 independent variable (Sperandei 2014)."

2 main goals for building a CCP are detecting customers with the highest propensity of churning and generating insights about the reasons why churners churn. This study aims to understand the drivers of churn; therefore, we used Random Forest and Logistic Regression algorithms because they are more suitable when it comes to comprehending the rationale behind the decision of the algorithm as well as revealing which variable is more effective in churn decision.

3.1.3 Model accuracy

Regardless of what classifier algorithm is selected, the model accuracy should be measured to see if the model is successful in predicting future data points. In a model with binary values in the target, the model makes the labeling with, say, 1 and 0, whereas the true values of the target are again 1 and 0. This creates a confusion matrix that shows the prediction success of the model, an example of which could be found in Figure 3.3. A successful model is one with high TP and low FP.

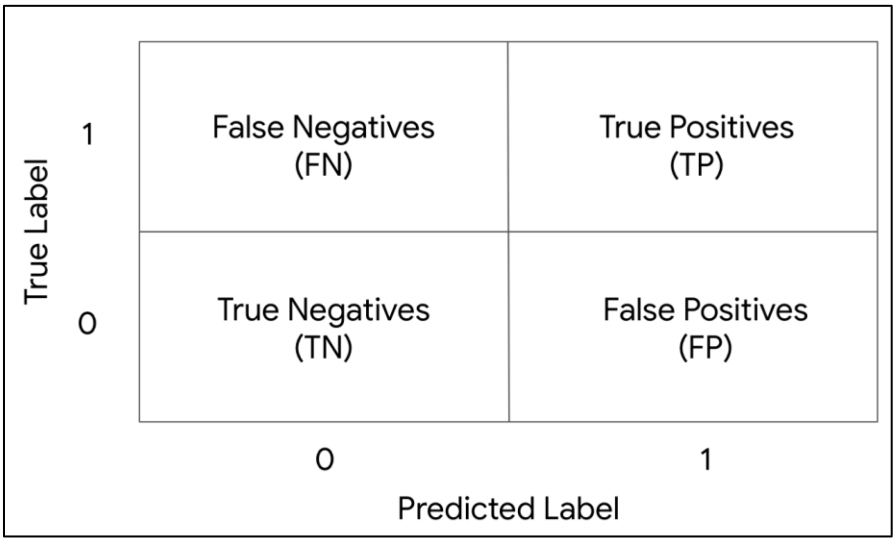

Figure 3.3 Confusion Matrix

There are also certain metrics derived from the confusion matrix such as precision, recall, F1 score, and accuracy score. The precision formula is: $\frac{True\ Positive}{True\ Positive+False\ Positive}$. It shows the model's capability of correctly labeling as positive. The recall formula is: $\frac{True\ Positive}{True\ Positive+False\ Negative}$. It shows the model's capability of catching all true positives. This is especially important in the datasets where the class imbalance is a problem where the model gets unproportionable and inclined to label as negative, which leads to a high false negative (FN) value. Accuracy score formula is: $\frac{TP+TN}{TP+FN+TN+FP}$. The accuracy score shows how accurate the model predicts and has been largely used in the machine learning field. However, it might be misleading in cases where class imbalance exists in the dataset. If the majority of the class is negative, the TN value will dominate the accuracy score and cause a false high accuracy score, which means precision is overvalued at the expense of the recall score.

F1 score formula is: $2 \times Precision\ Score \times \frac{Recall\ Score}{Precision\ Score+Recall\ Score}$
F1 score is more suitable when one does not want to sacrifice either precision or recall scores.

Another approach for measuring model accuracy is the Receive Operating Curve (ROC) and Area Under the ROC curve techniques (Bradley 1997, 1145-1159). The higher the AUC is, the more successful a model is.
ROC is drawn by the 2 metrics: True Positive Rate (TPR) and False Positive Rate (FPR). The TPR formula is: $\frac{TP}{TP + FP}$ and the FPR formula is: $\frac{FP}{FP + TN}$

3.1.4 Feature importance

Interpretability of a CCP is as important as the prediction accuracy of the model even if the model is solely built for prediction. Both classification and regression algorithms offer readable reports showing the importance of the features.

RF has variable importance and Gini index (impurity) techniques to display what features are of more importance. In the variable importance technique, features are permuted and the mean decrease in model accuracy is computed to determine feature importance. Gini Index technique measures how much a variable reduces the Gini Impurity metric in a particular class after each node split (Belgiu and Dragut 2016). The scikit-learn library sklearn.ensemble.RandomForestClassifier has a built-in method feature_importances_ that calculates the feature importances.

However, impurity-based calculations are vulnerable to high cardinality features and might therefore be misleading. Permutation importance is an alternative way to calculate feature importance and is also model-agnostic. In essence, a random subset of features is selected for n times and the model accuracy is calculated to see the importance of features. Eli5 library in Python allows this calculation. It is also possible to change the type of score for which features are permuted and the importance of them is calculated.

4. DATA

4.1. Data & Data Source

Data is obtained from a telecom company in an encrypted and privacy-safe way. The subscribers that are involved in the dataset are individual and postpaid. The main reason for excluding corporate subscribers is that the decision-making process about selecting the operator or even tariff plans is mostly not up to the end user. Certain agreements and contracts are made and the employees use the selected operator.

Moreover, the reason for excluding prepaid subscribers is the industry insight provided by the data sourcer that these two types of subscribers behave quite differently and their churn ratios also diverge significantly. This would be reflected in the usage-related metrics to be used in the churn model, causing an inferior performance.

Data includes 19 variables and 104561 entries in its raw format, i.e. before any pre-processing is applied. 16 of the variables are numeric and 3 of them are categorical. The dependent variable in the dataset is the churn status of subscribers in the selected months. It is reflected in the data as Target and 1 indicates that a subscriber has churned and 0 indicates that a subscriber has not churned.

As in many detection problems, data has class an imbalance problem. Of all the entries, there are 99992 subscribers who did not churn, and 4569 subscribers who churned in the selected months of the analysis. Some of the classification techniques are more impacted than others by the class imbalance problem. It is often regarded best to try out ways to handle the problem and see the differences. The class imbalance will be dug deeper later in this section.

Variables in this research could be summarized into 5 categories as customer statistics, customer satisfaction, OTT & digital channels usage metrics, telco usage metrics, and invoice details. This categorization would apply to all variables regardless of being numeric or categorical.

4.1.1. Numeric variables

Numeric variables are the average value of each variable in the 3 months (from t-1 to t-3) preceding the selected observation month (t) for churn status, in other words, target. The data summary and descriptive statistics summary can be seen in Table 4.1. It is worth noting that some of the variable names are masked in line with the data privacy concerns of the data sourcer.

Table 4.1 Descriptive Statistics

	# of entries	mean	std	min	25%	50%	75%	max
Age	103936	38.51	15.18	1	26	39	49	100
Tenure in Months	104561	103.14	79.31	3	34	87	144	331
# of Call Center Incoming Calls in Last 3 Months	33426	2.29	2.77	1	1	1	3	168
# of Complaints via Call Center	5348	1.26	0.69	1	1	1	1	13
Remaining Months Before Contract Expiration	101498	6.15	3.59	-12	3	6	9	14
Usage of OTT App1 (# of Transactions)	7647	38.83	60.6	0	3	15	50	844
Usage of OTT App2 (# of Transactions)	16874	168.06	195.15	0	27	98	234	999
Usage of OTT App3 (# of Transactions)	4498	4.72	25.62	0	1	1	3	876
# of Logins in Main Digital App	90622	12.03	15.31	0	3	8	15	434
# of Loyalty Giveaway Obtainments	81405	2.46	1.48	0	1	3	4	5
Voice Calls in Minutes	104561	320.71	292.84	0	10	265	540	999
# of SMS	104561	76.58	138.65	0	4	21	79	999
Data Usage in gigabytes	104561	44.9	38.33	0	19	36	59	658
Average Revenue per User (Turkish Liras)	103096	90.45	55.6	0	54	76	111	986
Excess Usage Revenue per User (Turkish Liras)	21587	27.75	30.85	1	8	16	33	531

Before digging deeper into the variables, it is also important to note that the missing ratio for some of the variables might seem strikingly high; however, the missing value in those variables means zero (0), based on the information received from the data sourcer. The data tables are structured in a way to fill null values if the subscriber does not have any record regarding that variable. Those variables where null value means zero (0) are: # of Call Center Incoming Calls in Last 3 Months, # of Complaints via Call Center, # of Loyalty Giveaway Obtainments, Excess Usage Revenue per User (Turkish Liras), # of Logins in Main Digital App, Usage of OTT App1 (# of Transactions), Usage of OTT App2 (# of Transactions), Usage of OTT App3 (# of Transactions).

Customer statistics: The variable Age falls under this category and its mean is 38.5. Its missing ratio is lower than 1% and could easily be handled with imputing techniques.

Customer satisfaction: Tenure in Months, # of Call Center Incoming Calls in Last 3 Months, # of Complaints via Call Center, Remaining Months Before Contract Expiration, # of Loyalty Giveaway Obtainments are the variables in this category.

33426 of the subscribers in our dataset had a call center record. Not every call center call originates from a complaining reason so there are 5,348 subscribers who contacted the call center for complaint purposes. One subscriber might have multiple complaints and call center calls, which is why the maximum value is bigger than 1 in these variables. Remaining Months Before Contract Expiration is how many months there are before the contract that the subscriber signed expires. # of Loyalty Giveaway Obtainments is the number of times a subscriber has received a loyalty giveaway product. Subscribers need to complete a couple of simple steps within the loyalty applications to earn these benefits.

OTT & Digital Channels Usage: Usage of OTT App1 (# of Transactions), Usage of OTT App2 (# of Transactions), Usage of OTT App3 (# of Transactions) and # of Logins in Main Digital App variables are the variables in this category. Each app is distinct but their names are masked for privacy concerns, as mentioned above.

Telco Usage: Voice Calls in Minutes, # of SMS, and Data Usage in Gigabytes are the variables in this category. These are commonly used variables among telecom operators such as voice, data and SMS usage stats.

Invoice Details: Average Revenue per User (Turkish Liras) and Excess Usage Revenue per User (Turkish Liras) are the variables in this category. Average revenue is more obvious than the other as its name is revealing. Excess Usage Revenue per User (Turkish Liras) is the revenue generated from a subscriber when he/she continues to consume after his/her tariff benefits are fully depleted.

4.1.2 Categorical variables

There are 3 categorical variables as Gender, Invoice Auto Payment Flag, and Contract Flag.

Customer statistics: The variable Gender falls under the 'Customer Statistics' category and it is a binary variable.

Invoice Details: Invoice Auto Payment Flag is whether a subscriber turned on automatization of paying bills to the operator. This one is also a binary variable with yes or no being the distinct values.

Customer satisfaction: Contract Flag is whether a subscriber has signed a contract with the operator. It was mentioned under the literature review part that contracts tend to create a switching barrier for subscribers; thus, decreasing the propensity to churn.

5. ANALYSIS

5.1 Data Pre-process

The target value in the telecom churn prediction model is the churn status of customers in the selected time period. By default, the churn status column has 2 values; ‘Churn’ or Null. Null means that a customer did not churn. As mentioned above, classifiers like the random forest can handle these types of data; however, we transform this feature as 1 and 0 (1 being churned, 0 being not churned) to get the best out of logistic regression because both classifiers will be used and compared to each other. The name of the column also has been changed from churn status to ‘Target’. We also checked for features with only 1 unique numeric value by filtering with standard deviation = 0. Some flag features were ‘1’ for all the records and we dropped them as well because the absence or presence of these features made no difference.

5.1.1 Transforming contract expiration month

A contract is a mutual agreement between a customer and a telecom operator on the details of the service offered to the customer. It usually outlines the duration, offers details (gigabyte, minutes, SMS, etc.), and the consequences if the customer decides to opt-out of the contract. By default, this variable is in the format of year and month, such as 2023-01, which is when the contract is to expire. We transformed this format in a way that it started showing how many months are left before the contract expires, like 5 or 7 as an example. There was also an arbitrary value that was assigned when the database failed to keep the actual record. We have changed that arbitrary value with null so that we could later impute it with certain imputing methodologies.

5.1.2 Handling missing values

The ratio of missing values in the records can be seen in Table 5.1. Of the methodologies for handling missing values, imputing with median is widely used. Therefore, we imputed, in other words, replaced, the missing values with the median of that variable, except for the Gender variable. The calculating median is not applicable for Gender as it is a binary variable with Female or Male unique records. Therefore, we eliminated those records where the variable Gender is missing.

Table 5.1 Missing Ratios in Features

	Missing Ratio
Gender	9.4%
Remaining Months Before Contract Expiration	2.9%
Average Revenue per User (Turkish Liras)	1.4%
Age	0.6%

5.1.3 Transforming categorical variables

There are 3 categorical variables in the data; Gender, Invoice Auto Payment Flag, and Contract Flag. Categorical variables are transformed into numeric variables by using the get_dummy function from the Python pandas library. This method creates a column for each unique value of the categorical feature that is transformed and then assigns 1 and 0 as a flag. For example, the feature Gender has 2 distinct values as Female and Male. After transformed by get_dummy function, Gender_M, which stands for Gender Male, is created. If Gender_M is 1, it means the gender is male. If Gender_M is 0, it means gender is female. There is no need to create an additional column as Gender_F in this case as Gender_M = 0 implies the subscriber is female.

5.1.4 Variable collinearity check

The multicollinearity heatmap that is shown in Figure 5.1 is utilized to visualize the correlation between the variables. In Figure 5.1, it is seen that there are 3 correlations whose coefficients are around 0.5. The first one is between # of Logins in Main Digital App and # of Loyalty Giveaway Obtainments. This correlation is expected since the loyalty benefit could be redeemed via the main digital channel. However, what the main digital channel of this telecom company offers involves more than only a loyalty giveaway. Also, the correlation coefficient is 0.59, which might not be evaluated as a strong correlation. Therefore, it will be insightful to keep these two variables in the model. The second one is between Data Usage in Gigabytes and Average Revenue per

User (Turkish Liras) and the correlation coefficient is 0.55. The third one is between Remaining Months Before Contract Expiration and Contract Flag and the correlation coefficient is 0.50. The coefficients are not strong, i.e., more than 0.80 so we leave the variables in the model.

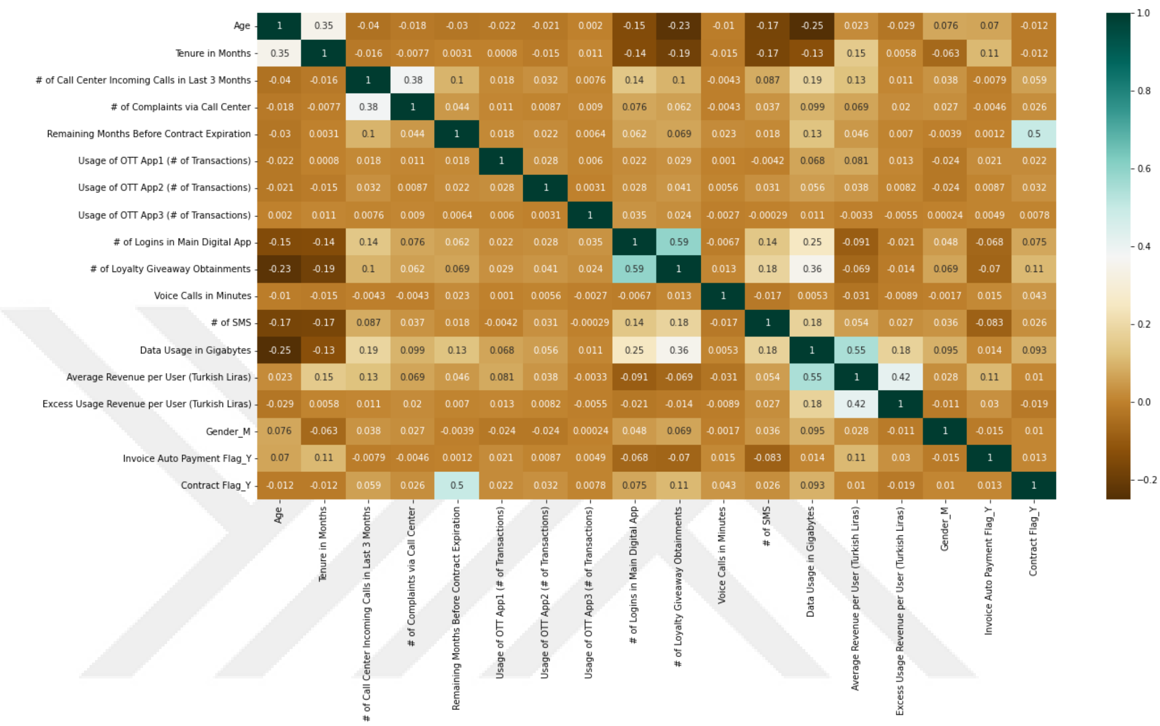

Figure 5.1 Correlation Analysis Results in a Heat Map

5.1.5 Scaling the variables

As in many datasets in the real world, the variables in our dataset have varying degrees of magnitudes. This creates an obstacle for some of the classification algorithms as they are sensitive to the scale of a variable. A variable whose values vary between 500 and 1000 should not be weighed in the predictive model more than another variable whose values are between 5 to 10 just because of the magnitude difference. For example, GPRS (internet data) usage of customers might be in the form of bytes, kilobytes, megabytes, or gigabytes. Although they mean the same thing but in different unit measures, in essence, the classification algorithms that are sensitive to different scales might evaluate this same GPRS usage variable differently in the model, depending on its unit measure.

Although tree-based classification algorithms are not highly sensitive to the scale of variables, the ones that use gradient descent-based algorithms to optimize the performance of a model require the variables to be scaled. Linear & Logistic Regression, Neural Networks fall within this category. Our dataset needs to be scaled because we will also utilize Logistic Regression in addition to the Random Forest classification algorithm. There are 2 different scaling techniques as standardization and normalization. We will use standardization. Standardization means a variable is centered around a mean value of zero with a standard deviation of one. The formula used in this process is as follows:

$$X' = \frac{X - \mu}{\sigma} \tag{5.1}$$

where μ is the mean and σ is the standard deviation of the values in the variable. We used sklearn.preprocessing, StandardScaler function from the sklearn library in python.

5.1.6 Class imbalance - smote algorithm to resample

In our dataset, the target is the churn status in the selected period. Class distribution is shown in Table 5.2 and there is a class imbalance in the dataset since there are 4.5K observations in class 1 while there are 90K in class 0.

Table 5.2 Class Distribution

TARGET	Class Count
0	90140
1	4545

As it can be seen from Figure 5.2, the model is much more inclined to label as 0 (negative), therefore; the number of true predicted 1 (positive) gets very low, which is the core reason why the model is built in the first place: maximizing the true positives. Metrics should be evaluated with care at this point since very high accuracy might just be because of the high number of true negatives as the machine learning algorithm is trained to predict the most common class in the dataset.

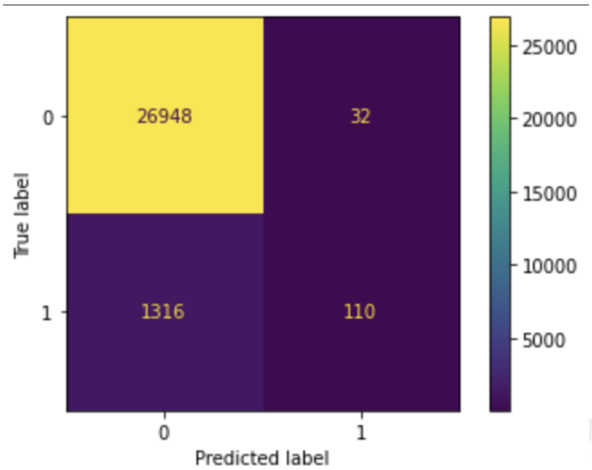

Figure 5.2 Confusion Matrix When Class Imbalanced

We applied the SMOTE algorithm via the imblearn python library (imblearn.over_sampling.smote). The shape of our data changed from (94685, 18) to (180280, 18) and the balance of classes changed to 50-50, meaning equal class distribution.

5.2 Test & Train Split

Now that pre-processing steps are complete, we can move on to fitting a classification algorithm. There are a few steps we need to take before the actual fitting action. Firstly, data needs to be divided into test and training subgroups to prevent overfitting and measure the results in a better way. The model will learn from the training and be ready to predict unseen future data. We used 25% as the test ratio meaning that 75% of the data is used as a training data set.

5.3 Prediction / Classification

The last step of building the CCP is fitting the classifier to the training dataset so that the model learns from known past data. We used Random Forest and Logistic Regression as the classifiers in this research.

6. RESULTS

Descriptive statistics results are shown in 4.1.1 Numeric Variables section and variable correlation results are shown in 5.1.5. Variable Collinearity Check section. Therefore, we directly display the model results.

6.1 The Performance of the Models

6.1.1 Random forest

We first look at the confusion matrix shown in Figure 6.1 to see how accurately Random Forest (RF) predicts. As we separated the 25% of our data to test split, there are 45070 data points in the test set. We also implemented the SMOTE algorithm to make the class distribution even between classes.

There are 22359 negatives, which means non-churners, in the test data set and RF successfully predicted 21801 as negatives, which is the True Negatives (TN). RF failed in 558 of the negatives by predicting them as positives, which is False Positives (FP).

There are 22711 positives, which means churners, in the test data set and RF successfully predicted 21377 as positives, which is True Positives (TP). RF failed in 1334 of the positives by predicting them as negatives, which is False Negatives (FN).
The accuracy score of the model is 0.96 and F-1 score is also 0.96, which indicates that the model is very strong.

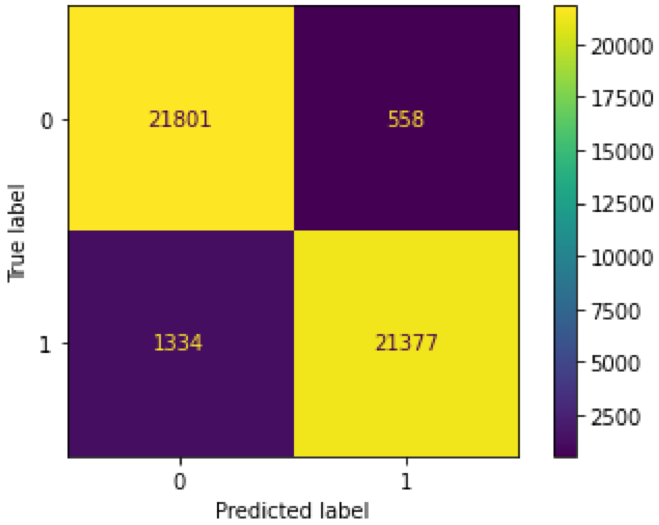

Figure 6.1 The confusion matrix of Random Forest

The ROC curve shown in Figure 6.2 is drawn using the TPR and FPR metrics and the corresponding AUC score is 0.99.

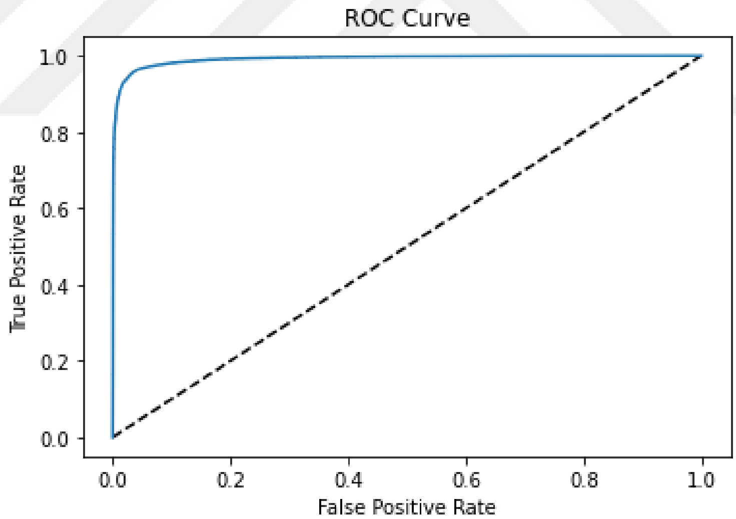

Figure 6.2 ROC curve of Random Forest

6.1.2 Logistic regression

The confusion matrix is shown in Figure 6.3 to display how accurately Logistic Regression (LR) predicts.

There are 22359 negatives in the test data set and LR successfully predicted 15379 as negatives, which is the True Negatives (TN). LR failed in 6980 of the negatives by predicting them as positives, which is False Positives (FP).

There are 22711 positives, which means churners, in the test data set and LR successfully predicted 15586 as positives, which is True Positives (TP). LR failed in 7125 of the positives by predicting them as negatives, which is False Negatives (FN).

The accuracy score of the model is 0.69 and F-1 score is also 0.69, which indicates that the model is not strong and its performance is worse than that of Random Forest.

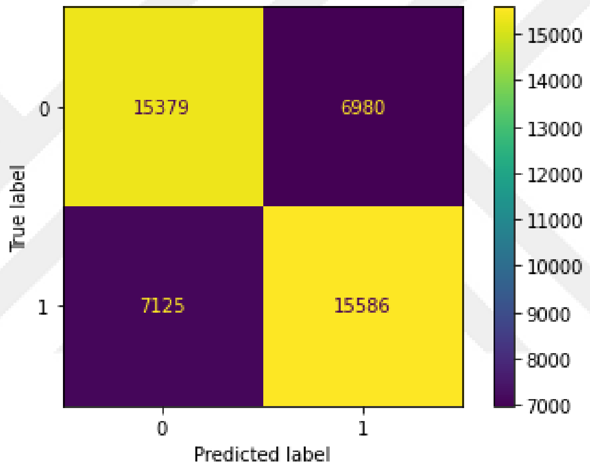

Figure 6.3 The confusion matrix of Logistic Regression

The ROC curve shown in Figure 6.4 is drawn using the TPR and FPR metrics and the corresponding AUC score is 0.76.

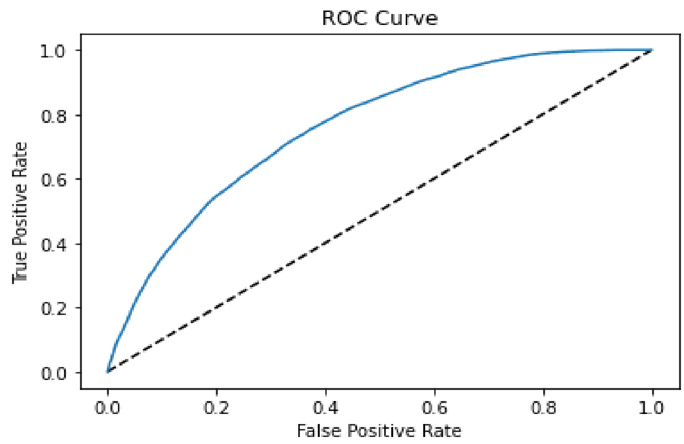

Figure 6.4 ROC curve of Logistic Regression

Therefore, our research proves that Random Forest is a better classifier for predicting churn with this data set.

6.2 Feature Importances

To determine what factors have an impact on the churn decision of customers, we look at feature importance in the churn prediction model. Since it is concluded that Random Forest is a better classifier for this model, we will only consider feature importance techniques for Random Forest.

The first method is the mean decrease in impurity (Gini). The mean decrease in impurity (MDI) metric measures the homogeneity of the nodes after each split by a variable. The higher the mean decrease in impurity, the more important the feature is. Figure 6.5 displays the results of feature importance using this technique. The figure claims that Remaining Months Before Contract Expiration, Age, and # of Loyalty Giveaway Obtainments are the 3 most important features in the churn decision of customers. # of Logins in Main Digital App, # of Call Center Incoming Calls In Last 3 Months and Voice Calls in Minutes are the following 3 most important features. The next most important features are also telecom-specific metrics such as Tenure in Months, # of SMS, Average Revenue per User (Turkish Liras), Data Usage in Gigabytes, and Excess Usage Revenue per User (Turkish Liras), which indicates that the churn decision of customers relates mostly to telecom-specific metrics.

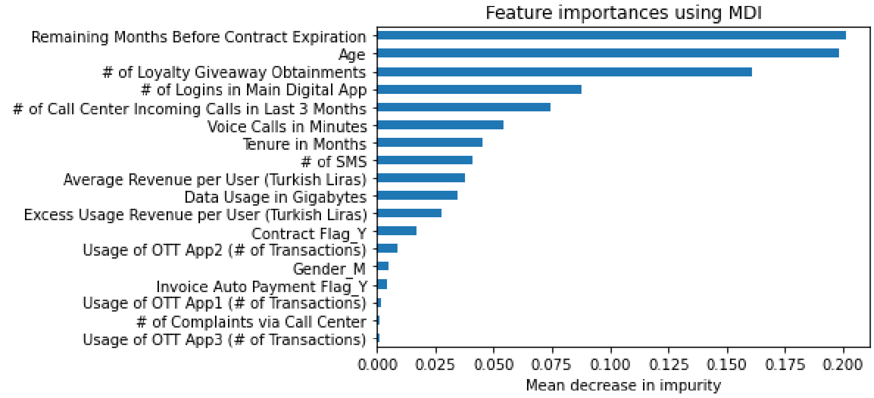

Figure 6.5 Feature Importance Results using MDI

The second method is the permutation importance. In this method, each feature is permuted and the mean decrease in the model accuracy score is calculated to determine how important a feature is. Since MDI might be misleading in favor of high cardinality features, it is important to see if and how feature importance is changing by this method.

It can be seen from Figure 6.6 that there is a slight change in the rankings of features by their importance as compared to the MDI technique. Although the 3 most important features are the same, the first one is now # of Loyalty Giveaway Obtainments, the second is Remaining Months Before Contract Expiration and the third is Age. The next 3 most important features are # of Call Center Incoming Calls in Last 3 Months, # of Logins in Main Digital App and Voice Calls in Minutes. There is no change in ranking in the next most important features, which are telecom metrics.

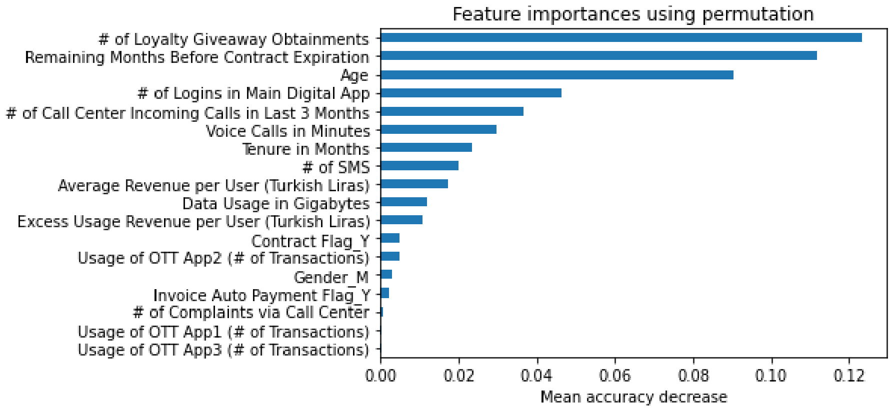

Figure 6.6 Feature Importance Results using Permutation

7. CONCLUSION

7.1 Discussion

This research study provides insights into the determinants of churn decisions of the customer in the telecom industry. It examines the effectiveness of not only the traditional telecom features such as gprs data usage, and invoice amounts but also digital loyalty apps and OTT media services. In addition to individual impacts of these features, the results also provide insights regarding the comparable importance of the features.

The findings indicate that the main digital channel, within which loyalty programs are to be found, and the usage of loyalty giveaways are significant factors affecting the churn decision. However, the usage of OTT media services plays an unimportant role in the churn decision of customers. Moreover, our findings also indicate that Random Forest is a better classifier than Logistic Regression for the specific data used in this study. There is no evidence; however, to generalize this result because the performance of classifiers may vary depending on the nature of the data.

As it has been put forth in the relationship marketing theory, retaining the existing customers and strengthening bonds between customers and brands are the foundations of sustainable competitive power. This study proves that loyalty programs and loyalty apps that strengthen the relationship between customer and brand are much more powerful than the strategy of product differentiation, which is over-the-top media service offerings in our case.

However, this study does not contain or reveal any insights as to why OTT media services are deemed insignificant in churn decisions by customers. Therefore, further qualitative studies are required to deep dive into the reasons for this decision. Yet, it can be suggested that customers perceive the offerings of loyalty programs as more valuable as well as more comparable to that of others concerning churn decisions. Additionally, this study contains data from a Turkish telecom company, which means the preferences and actions of customers might vary depending on the geography.

To sum up the research findings, below are the answers to the research questions and hypotheses:

Research question 1: Do OTT apps have a significant impact on the churn decision of customers?
Answer: No, OTT apps do not have a significant impact on the churn decision of customers.

Research question 2: Does the usage of main digital channels and loyalty apps have a significant impact on the churn decision of customers?
Answer: Yes, the usage of the main digital channel, which also has loyalty apps within, is of high importance in the churn model; therefore, it has a significant impact on the churn decision of customers.

Research question 3: Do loyalty giveaways have a significant impact on the churn decision of customers?
Answer: Yes, in fact, obtainment of loyalty giveaway is the highest important feature in the model according to feature importance with the permutation method.

Research question 4: What features have the most impact on the churn decision of customers?
Answer: The top 5 features based on highest importance in the churn model are # of Loyalty Giveaway Obtainments, Remaining Months Before Contract Expiration, Age, # of Call Center Incoming Calls in Last 3 Months, and # of Logins in Main Digital App. The top 5 features do not change regardless of the methodology of feature importance calculation; however, their ranking changes. Yet, as stated in the analysis section, we prefer permutation importance to MDI methodology because of certain drawbacks pertaining to MDI methodology.

7.2 Conclusion

This research aims to examine if the digital strategies applied in the telecom industry, which involves OTT media apps and loyalty apps, play an effective role in the churn decision of customers.

A customer churn prediction model is built to see which factors are the most important in customer churn decisions. Random Forest and Logistic Regression classification algorithms are used and their predictive power results are compared. Random Forest yielded better performance. Based on the feature importance analysis of the churn prediction model, it is seen that loyalty apps and loyalty giveaway program are very effective while OTT apps play an unimportant role in the churn decision of customers.

Although the feature importance methodology reveals what factors are important in the churn prediction model, it does not uncap root causes as to why customers give higher or lower importance to the features. The successful effectiveness of the loyalty app and loyalty giveaway program aligns with the relationship marketing theory, which claims that a closer relationship with the customer and personalized programs tend to yield lower churns as a result of a stronger bond between customer and brand.

The result that OTT media apps play little role in the churn decision of customers does not mean that this strategy adds no value to the overall strategy of the brand. It is clear that OTT media apps constitute an indirect but strong threat to telecom operators in the long run. Therefore, OTT strategies might be better positioned as a long-term transformation of the core telecom business as opposed to being a booster of the current offerings. This study could be further improved by analyzing data from different telecom companies and regions. Not only can the customer behavior in a specific country change, but also a different telecom company in the same country might position its digital strategies differently; thus, causing a different effect on the churn decision of its customers. Moreover, as in most of the industries, seasonality might also lead to a certain degree of changes in the outcome of this study. Therefore, a longer or different period of time selection has the potential to result in different inferences.

REFERENCES

Aaker, D. A. 1991. *Managing Brand Equity*. New York: The Free Press.

Ahn, J. H., Han, S. P., Lee, Y. S. 2006. “Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry.” *Telecommunications Policy* 30(10-11):552-568. https://doi.org/10.1016/j.telpol.2006.09.006

Amin, A., Al-Obeidat, F., Shah, B., Tae, M., Khan, C., Durrani, H., and Anwar, S. 2017. “Just-in-time customer churn prediction in the telecommunication sector.” *Journal of Supercomputing*. https://doi.org/10.1007/s11227-017-2149-9

Athanassopoulos, A. 2000. “Customer satisfaction cues to support market segmentation and explain switching behavior.” *Journal of Business Research* 47 (3):191–207. https://doi.org/10.1016/s0148-2963(98)00060-5

Attenborough, N., J. Sandbach, U. Saadat, G. Siolis, M. Cartwright, and S. Dunkley 1998. "Feasibility Study and Cost Benefit Analysis of Number Portability for Mobile Services in Hong Kong." *Final Report for OFTA prepared by NERA and Smith System Engineering*.

Belgiu, M., and Drăguţ, L. 2016. “Random forest in remote sensing: A review of applications and future directions.” *ISPRS Journal of Photogrammetry and Remote Sensing* 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Boulding, W., Staelin, R., Ehret, M., and Johnston, W.J. 2005. “ A Customer Relationship Management Roadmap: What is Known, Potential Pitfalls and Where to Go.” *Journal of Marketing* 69 (4):155–166. https://doi.org/10.1509/jmkg.2005.69.4.155

Bradley, A. P. 1997. “The use of the area under the roc curve in the evaluation of machine learning algorithms.” *Pattern Recognition* 30(7):1145–1159. https://doi.org/10.1016/s0031-3203(96)00142-2

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45(1):5–32. https://doi.org/10.1023/a:1010933404324

BTK (Bilişim Teknolojileri Kurumu) 2021. *Quarterly Market Report.* https://www.btk.gov.tr/uploads/pages/pazar-verileri/ceyrek-raporu-2021-2-ceyrek-30-09-21-v8-kurumdisi.pdf

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. 2002. “SMOTE: Synthetic Minority Over-sampling Technique.” *Journal of artificial intelligence research* 16:321-357. https://doi.org/10.1613/jair.953

Confraria, J., Ribeiro, T., and Vasconcelos, H. 2017. “Analysis of consumer preferences for mobile telecom plans using a discrete choice experiment.” *Telecommunications Policy* 41(3):157–169. https://doi.org/10.1016/j.telpol.2016.12.009

Das, K. 2009. "Relationship marketing research (1994-2006): An academic literature review and classification." *Marketing Intelligence & Planning* 27(3):326-363. https://doi.org/10.1108/02634500910955236

Datta, P.R. 2003. “The determinants of brand loyalty.” *Journal of American Academy of Business* 3(1/2):138-144.

Durukan, T., Bozacı, İ., and Doğan, T. T. 2011. "Mobile number portability in Turkey: An empirical analysis of consumer switching behavior." *European Journal of Social Sciences* 20(4): 572-585. https://www.researchgate.net/publication/287000390

Faircloth, J. B., Capella, L. M., and Alford, B. L. 2001. “The Effect of Brand Attitude and Brand Image on Brand Equity.” *Journal of Marketing Theory and Practice* 9(3):61–75. https://doi.org/10.1080/10696679.2001.11501897

Farooq, M., and Raju, V. 2019. “Impact of Over-the-Top (OTT) Services on the Telecom Companies in the Era of Transformative Marketing.” *Global Journal of Flexible Systems Management* 20(2):177–188. https://doi.org/10.1007/s40171-019-00209-6

Ferguson, R. and Brohaugh, B. 2008. "Telecom's search for the ultimate customer loyalty platform." *Journal of Consumer Marketing* 25(5):314-318. https://doi.org/10.1108/07363760810890543

Frow, P. E., and A. F. Payne. 2009. “Customer Relationship Management: A Strategic Perspective.” *Journal of Business Market Management* 3(1):7–27. https://doi.org/10.1007/s12087-008-0035-8

Ganesh, J., Arnold, M. J., & Reynolds, K. E. 2000. “Understanding the customer base of service providers: An examination of the differences between switchers and stayers.” *Journal of Marketing* 64(3):65–87. https://doi.org/10.2307/3203488

Ganuza, J. and Viecens, M. F. 2014. "Over-the-top (OTT) content: implications and best response strategies of traditional telecom operators. Evidence from Latin America." *Info* 16(5):59-69. https://doi.org/10.1108/info-05-2014-0022

Gerpott, T. J., Rams, W., and Schindler, A. 2001. “Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market.” *Telecommunications Policy,* 25(4):249–269. https://doi:10.1016/s0308-5961(00)00097-5

Grabner-Kraeuter, S. and Moedritscher, G. 2002. “Alternative Approaches Toward Measuring CRM Performance.” Sixth Research Conference on Relationship Marketing and Customer Relationship Management, Atlanta, June 9–12, 2002.

Green, W. and Lancaster, B. 2006. “Over the Top Services.” *LTC International* 7(4): 2.

Gronholdt, L., Martensen, A., and Kristensen, K. 2000. “The relationship between customer satisfaction and loyalty: Cross-industry differences.” *Total Quality Management* 11(4–6):509–514. https://doi.org/10.1080/09544120050007823

Grönroos, C. 1994. “From Marketing Mix to Relationship Marketing: Towards a Paradigm Shift in Marketing.” *Asia-Australia Marketing Journal* 2(1):9–29. https://doi.org/10.1016/s1320-1646(94)70275-6

Gupta, A., and Sahu, G. P. 2015, “Exploring Relationship Marketing Dimensions and Their Effect on Customer Loyalty- A Study of Indian Mobile Telecom Market.” *International Journal of Business Innovation and Research* 9(4):375-390. https://doi.org/10.1504/ijbir.2015.070176

Gustafsson, A., Roos, I., and Edvardsson, B. 2004. "Customer clubs in a relationship perspective: a telecom case." *Managing Service Quality: An International Journal* 14(2/3):157-168. https://doi.org/10.1108/09604520410528581

Ho, Tin Kam. 1995. "Random decision forests." *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1:278–282. https://doi.org/10.1109/ICDAR.1995.598994

Howitt, D., and Cramer, D. 2011. An introduction to statistics in psychology (5th ed.). Pearson.

Huang, B., Kechadi, M. T., and Buckley, B. 2012. ”Customer churn prediction in telecommunications.” *Expert Systems with Applications* 39(1):1414-1425. https://doi.org/10.1016/j.eswa.2011.08.024.

Huang, J., Li, Y. F., and Xie, M. 2015. “An empirical analysis of data preprocessing for machine learning-based software cost estimation.” *Information and Software Technology* 67:108–127. https://doi.org/10.1016/j.infsof.2015.07.004

Hwang, H., Jung, T., and Suh, E. 2004. “An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry.” *Expert Systems with Applications* 26:181–188. https://doi.org/10.1016/s0957-4174(03)00133-7

Ishak, F., and Abd. Ghani, N. H. 2013. “A Review of the Literature on Brand Loyalty and Customer Loyalty.” *In: Conference on Business Management Research, Sintok, December 11, 2013*:186-198. http://repo.uum.edu.my/16316/1/20.pdf

Jacoby, J. and Chestnut, R.W. 1978. *Brand Loyalty: Measurement and Management*. New York: John Wiley & Sons.

Kale, S. H. 2003. “CRM in Gaming: It’s No Crapshoot!” *UNLV Gaming Research & Review Journal* 7(2):4. https://digitalscholarship.unlv.edu/grrj/vol7/iss2/4/

Keaveney, S. M. 1995. "Customer switching behavior in service industries: An exploratory study." *Journal of Marketing* 59(2):71-82. https://doi.org/10.1177/002224299505900206

Keller, K. L. 1993. "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity." *Journal of Marketing* 57(1):1-22. https://doi.org/10.2307/1252054

Kim, H. S., and Kwon, N. 2003. “The advantage of network size in acquiring new subscribers: a conditional logit analysis of the Korean mobile telephony market.” *Information Economics and Policy* 15(1):17–33. https://doi.org/10.1016/s0167-6245(02)00070-7

Kim, H. S., and Yoon, C. H. 2004. “Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market.” *Telecommunications Policy* 28(9–10):751-765. https://doi.org/10.1016/j.telpol.2004.05.013.

Kim, M. K., Park, M. C., and Jeong, D. H. 2004. “The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services.” *Telecommunications Policy* 28(2):145–159. https://doi.org/10.1016/j.telpol.2003.12.003

Kisioglu, P., and Topcu, Y. I. 2011. “Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey.” *Expert Systems with Applications* 38(6):7151-7157. https://doi.org/10.1016/j.eswa.2010.12.045

Kotler, P., & Armstrong, G. 2003. *Principles of marketing*. New Jersey: Prentice Hall PTR.

Kotler, P., Dubois, B., and Manceau, D. 2003. *Marketing Management, 11th ed*. Prentice-Hall: Upper Saddle River, NJ. https://www.pearson.ch/download/media/9782744070402_toc.pdf

Kotsiantis, S. B., Kanellopoulos, D., and Pintelas, P. E. 2006. "Data preprocessing for supervised learning." *International journal of computer science* 1(2): 111-117.

Lee, J., Lee, J., and Feick, L. 2001. "The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France." *Journal of Services Marketing* 15(1):35–48. https://doi.org/10.1108/08876040110381463

Liu, Z., Hardy, J., and Zhang, J. 2019. "A practical study on imbalanced data re-sampling for conversion rate of online advertising." *Turkish Journal of Electrical Engineering and Computer Sciences* https://www.researchgate.net/publication/330742123

Madden, G., Savage, S. J., and Coble-Neal, G. 1999. "Subscriber churn in the Australian ISP market" *Information Economics and Policy* (11/2):195–207. https://doi.org/10.1016/S0167-6245(99)00015-3

Mascarenhas, O. A., Kesavan, R., and Bernacchi, M. 2006. "Lasting customer loyalty: a total customer experience approach." *Journal of Consumer Marketing* 23(7):397–405. https://doi.org/10.1108/07363760610712939

Mellens, M., DeKimpe, M. G., and Steenkamp, E. M. 1996. "A Review of Brand-Loyalty Measures in Marketing." *Journal of Economic Management* 41(4) :507-533.

Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E., and Kaushansky, H. 2000. "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry." *IEEE Transactions on Neural Networks* 11(3):690–696. https://doi.org/10.1109/72.846740

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., and Mason, C. H. 2006. "Defection detection: measuring and understanding the predictive accuracy of customer churn models." *Journal of Marketing Research* 43(2):204–211. https://doi.org/10.1509/jmkr.43.2.204

Nikou, S., Bouwman, H., De Reuver, M. 2012. "The potential of converged mobile telecommunication services: a conjoint analysis." *Info* 14(5): 21–35. https://doi.org/10.1108/14636691211256287

Oliver, R.L. 1997. Satisfaction: A Behavioral perspective on the consumer. New York: McGraw Hill.

Oskarsdottir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., and Vanthienen, J. 2016. "A comparative study of social network classifiers for predicting churn in the telecommunication industry." *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*:1151 - 1158. https://doi.org/10.1109/asonam.2016.775238

Parvatiyar, A., and Sheth, J. N. 2001. Conceptual Framework of Customer Relationship Management: Emerging Concepts, Tools and Applications. New Delhi, India: Tata McGraw-Hill:3–25.

Peppers, D., and Rogers, M. 1993. The One to One Future: Building Relationships One Customer at a Time. New York: Currency Doubleday.

Peppers, D., and Rogers, M. 1999. Enterprise One-to-One: Tools for Competing in the Interactive Age. New York: Doubleday.

PWC 2017. *An industry at risk: Commoditization in the wireless telecom industry.* https://www.strategyand.pwc.com/gx/en/insights/2017/industry-at-risk.html

Reichheld, F. F., and Sasser. W. E. 1990. "Zero defeofions: Quoliiy comes to services." *Harvard Business Review* 68(5):105–111. http://matlesiouxx.free.fr/Cours/HKU/Courses/CSIS0404/Lecture%201/Module1_3_Zero_Defections_-Quality_Comes_to_Services.pdf

Richter, Y., and Slonim, N. 2010. "Predicting customer churn in mobile networks through analysis of social groups." *Proceedings of the 2010 SIAM International Conference on Data Mining*:732–741. https://doi.org/10.1137/1.9781611972801.64

Ryals, L., and Payne, A. 2001. "Customer relationship management in financial services: towards information-enabled relationship marketing." *Journal of Strategic Marketing* 9(1): 3–27. https://doi.org/10.1080/713775725

Saarela, M., and Jauhiainen, S. 2021. "Comparison of feature importance measures as explanations for classification models." *SN Applied Science* 3: 272. https://doi.org/10.1007/s42452-021-04148-9

Schneider, Benjamin. 1980. "The Service Organization: Climate is Crucial." *Organizational Dynamics* 9(2): 52-65. https://doi.org/10.1016/0090-2616(80)90040-6

Selland, C. and Pockard, D. 2003. "Are Companies Responsible for their CRM Failures?" *Network World* 20(28):29.

Shin, D. H. 2012. "What makes consumers use VoIP over mobile phones? Free riding or consumerization of new service." *Telecommunications Policy* 36(4):311-323. https://doi.org/10.1016/j.telpol.2012.01.004

Sperandei, S. 2014. "Understanding logistic regression analysis." *Biochemia Medica:12–18.* https://doi.org/10.11613/bm.2014.003

Sujata, J., Sohag, S., Tanu, D., Chintan, D., Shubham, P., and Sumit, G. 2015. "Impact of Over the Top (OTT) Services on Telecom Service Providers." *Indian Journal of Science and Technology* 8(S4): 145–160. https://doi.org/10.17485/ijst/2015/v8iS4/62238

Thakur, R., Summey, J. H. and Balasubramanian, S.K. 2006. “CRM as Strategy: Avoiding the Pitfall of Tactics.” *Marketing Management Journal* 16 (2):147–154.

Uncles, M. D., Dowling, G. R., and Hammond, K. 2003. “Customer loyalty and customer loyalty programs.” *Journal of Consumer Marketing* 20(4):294–316. https://doi.org/10.1108/07363760310483676

Uner, M. M., Guven, F., and Cavusgil, S. T. 2020. “Churn and loyalty behavior of Turkish digital natives: Empirical insights and managerial implications.” *Telecommunications Policy* 44(4): 101901. https://doi.org/10.1016/j.telpol.2019.101901

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., and Baesens, B. 2012. “New insights into churn prediction in the telecommunication sector: A profit driven data mining approach.” *European journal of operational research* 218(1):211-229. https://doi.org/10.1016/j.ejor.2011.09.031

Verhoef, P. C. 2003. “Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development.” *Journal of Marketing* 67(4):30–45. https://doi.org/10.1509/jmkg.67.4.30.18685

Zhao, M., Zeng, Q., Chang, M., Tong, Q., and Su, J. 2021. “A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China.” *Discrete Dynamics in Nature and Society*. https://doi.org/10.1155/2021/7160527

CURRICULUM VITAE

Personal Information
Name Surname : Ömer Buğra Kırgız

Education
Undergraduate Education : Bogazici University Management
Graduate Education : Kadir Has University Business Intelligence and Analytics
Foreign Language Skills : Turkish and English

Work Experience
Name of Employer and Employment Dates: Turkcell 2016-2020, Google 2020-Present