

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**PREDICTION OF CHURN RATE FOR PREPAID CUSTOMERS IN TURKISH
TELECOM INDUSTRY**



M.Sc. THESIS

Cansu ALATAS

Department of Management Engineering

Management Engineering Programme

JUNE 2018

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

**PREDICTION OF CHURN RATE FOR PREPAID CUSTOMERS IN TURKISH
TELECOM INDUSTRY**



M.Sc. THESIS

Cansu ALATAS
(507141008)

Department of Management Engineering

Management Engineering Programme

Thesis Advisor: Assoc. Dr. Elif KARAOSMANOGLU

JUNE 2018

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**TELEKOM SEKTÖRÜNDE FATURASIZ ABONELER İÇİN ABONE
HAREKETİ TAHMİNLEMESİ**

YÜKSEK LİSANS TEZİ

**Cansu ALATAŞ
(507141008)**

İşletme Mühendisliği Anabilim Dalı

İşletme Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Elif KARAOSMANOĞLU

HAZİRAN 2018

Cansu Alatas, a **M.Sc student** of **ITU Graduate School of Science Engineering and Technology** student ID 507141008, successfully defended the thesis/dissertation entitled “**PREDICTION OF CHURN RATE FOR PREPAID CUSTOMERS IN TURKISH TELECOM INDUSTRY**”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Doc. Dr. Elif KARAOSMANOĞLU**
Istanbul Technical University

Jury Members : **Assoc. Prof. Dr. Elif KARAOSMANOĞLU**
Istanbul Technical University

Prof. Dr. Burç ÜLENGİN
Istanbul Technical University

Dr. Lecturer Gülberk Gültekin SALMAN
Bahcesehir University

Date of Submission : 4 May 2018
Date of Defense : 30 May 2018



FOREWORD

This thesis is written as completion to the master of Management Engineering, at the Istanbul Technical University.

In this thesis, prepaid customers' behavioral and demographic data is investigated and churn rate probability is calculated using Decision Tree method.

I would like to thank to my advisors, **Assoc. Prof. Dr. Elif KARAOSMANOĞLU** and **Prof. Dr. Burc ÜLENGİN** especially for their excellent guidance and support during this process. I also would like to thank to my family, friends, especially my colleagues in my company for their cooperation.

June 2018

Cansu ALATAS



TABLE OF CONTENTS

| | <u>Page</u> |
|--|-------------|
| FOREWORD | vii |
| TABLE OF CONTENTS | ix |
| ABBREVIATIONS | xi |
| LIST OF TABLES | xiii |
| LIST OF FIGURES | xv |
| SUMMARY | xvii |
| ÖZET | xix |
| 1. INTRODUCTION | 1 |
| 2. OVERVIEW TO TELECOMMUNICATION SECTOR | 3 |
| 2.1 Evolution of the Telecommunication Market..... | 3 |
| 2.2 Turkey Market Statistics..... | 6 |
| 3. CUSTOMER CHURN | 7 |
| 3.1 Behavioral Loyalty | 13 |
| 3.2 Churn in Telecom Industry..... | 15 |
| 3.3 Churn Management Methods | 15 |
| 4. RESEARCH DESIGN | 17 |
| 4.1 Aim and Scope of the Research..... | 17 |
| 4.2 Data Preparation and Sampling | 17 |
| 4.2.1 Dataset interpretation | 19 |
| 4.2.2 Relation of variables..... | 23 |
| 4.3 Churn Prediction Methods..... | 38 |
| 4.3.1 Artificial neural network | 38 |
| 4.3.2 Support vector machine..... | 39 |
| 4.3.3 Naïve bayes | 39 |
| 4.3.4 Decision tree algorithm | 39 |
| 5. ANALYSIS AND FINDINGS | 41 |
| 5.1 Implementation..... | 41 |
| 5.2 Interpretation of the Findings | 44 |
| 5.3 Limitation of the Research | 45 |
| 6. CONCLUSION | 47 |
| REFERENCES | 49 |
| CURRICULUM VITAE | 55 |



ABBREVIATIONS

CAGR : Compound Annual Growth Rate

CAPEX : Capital Expenditures

IoT : Internet of Things

M&A : Merge and Acquisition

MNP : Mobile Number Portability

MoU : Minutes of Usage

OOB : Out of Bundle

OTT : Over the Top

VAS : Value Added Services



LIST OF TABLES

| | <u>Page</u> |
|---|-------------|
| Table 3.1: Previous studies about churn..... | 11 |
| Table 4.1: Training dataset with numbers..... | 18 |
| Table 4.2: Definition of the selected variables..... | 18 |
| Table 4.3: Offnet usage rate similarity..... | 19 |
| Table 4.4: OOB usage minutes similarity..... | 19 |
| Table 4.5: Tenure similarity..... | 19 |
| Table 4.6: Age similarity..... | 20 |
| Table 4.7: Top-up similarity..... | 20 |
| Table 4.8: Rate of incoming/ outgoing calls similarity..... | 20 |
| Table 4.9: Duration and SMS usage rate similarity..... | 21 |
| Table 4.10: Data usage rate similarity..... | 21 |
| Table 4.11: # of Migration similarity..... | 21 |
| Table 4.12: # of Port-in and Port-out similarity..... | 22 |
| Table 4.13: VAS product rate similarity..... | 22 |
| Table 4.14: OOB Data and SMS usage similarity..... | 22 |
| Table 4.16: Age-churn status Chi-square results..... | 24 |
| Table 4.15: Age-churn status test..... | 23 |
| Table 4.18: OOB usage-churn status Chi-square results..... | 25 |
| Table 4.17: OOB usage-churn status test..... | 24 |
| Table 4.20: Offnet usage rate-churn status Chi-square results..... | 26 |
| Table 4.19: Offnet usage rate-churn status test..... | 25 |
| Table 4.21: Incoming/outgoing rate-churn status test..... | 26 |
| Table 4.22: Incoming/outgoing rate-churn status Chi-square results..... | 27 |
| Table 4.23: Duration usage rate-churn status test..... | 27 |
| Table 4.24: Duration usage rate-churn status Chi-square results..... | 28 |
| Table 4.25: Data usage rate-churn status test..... | 28 |
| Table 4.26: Data usage rate-churn status Chi-square results..... | 29 |
| Table 4.27: SMS usage rate-churn status test..... | 29 |
| Table 4.28: SMS usage rate-churn status Chi-square results..... | 30 |
| Table 4.29: Tenure -churn status test..... | 30 |
| Table 4.30: Tenure-churn status Chi-square results..... | 31 |
| Table 4.31: Migration -churn status test..... | 31 |
| Table 4.32: Migration-churn status Chi-square results..... | 32 |
| Table 4.33: OOB Data usage - OOB SMS usage test using non-churner dataset..... | 32 |
| Table 4.34: OOB Data usage rate- OOB SMS usage rate results using non-churner dataset..... | 33 |
| Table 4.35: Port-in- port-out test using non-churner dataset..... | 33 |
| Table 4.36: Port-in- port-out results using non-churner dataset..... | 34 |
| Table 4.37: OOB data usage- top-up test using non-churner dataset..... | 34 |
| Table 4.38: OOB data usage rate- top-up results using non-churner dataset..... | 35 |

Table 4.39: OOB Data usage- OOB SMS usage test using churner dataset **35**
Table 4.40: OOB Data usage- OOB SMS usage results using churner dataset **36**
Table 4.41: Port-in- port-out test using churner dataset **36**
Table 4.42: Port-in- port-out results using churner dataset **37**
Table 4.43: OOB data usage rate- top-up test using churner dataset **37**
Table 4.44: OOB data usage rate- top-up results using churner dataset **38**
Table 5.1: First churn probability chart **42**
Table 5.2: Second churn probability chart **42**



LIST OF FIGURES

| | <u>Page</u> |
|---|-------------|
| Figure 2.1: Main growth trends in telecom in 1970-2003. | 4 |
| Figure 3.1: Phases of churn. | 12 |
| Figure 5.1: First result of churn probability. | 43 |
| Figure 5.2: Second result of churn probability. | 44 |





PREDICTION OF CHURN RATE FOR PREPAID CUSTOMERS IN TURKISH TELECOM INDUSTRY

SUMMARY

Telecommunication sector is the one of the competitive markets in business world. It is evolved from telegraph to mobile services in a short span of time due to technologic developments. Therefore, competition between telecommunication companies has increased rapidly over time.

As managing the competition among competitors, companies should know their customer behaviors and act preventively. In real world, it is hard to predict customers' behaviors and hence taking actions regarding each customer in companies' bases. Therefore, actions should be reduced and be focused on a specific group of customers. Also, it is known that new acquisition cost of the customer is much more higher than keeping the base (existing) customer not only in telecommunication but also in other industries. On the other hand, customer churn becomes easy with Mobile Number Portability (MNP) capability in the telecom industry. However, retention activities for churning customers are costly too. At this point, customer churn methods help companies to predict customers' brand switching behaviours to the other operators.

Prediction of the churn provides detection of potential churners and helps to identify the strategies to keep them by the companies. Some algorithms and methods are offered in previous research for the calculation of churn possibility.

In this thesis, it is aimed that, customer churn in telecom industry can be predicted with high degree accuracy using customer behavior dataset and decision tree method. This method is implemented on SPSS V22.0 program and the results are analyzed. In conclusion, behavioral and demographic dataset lead to churn rate calculation. Especially, offnet usage minutes which means call to other operators, out of bundle (OOB) usage minutes which means exceeded tariff minutes, customer tenure which means stayed months in customer base and customer age are the most accurate variables for predicting prepaid customer churn.

The main limitation of the analysis lies in the issue that prepaid customers are churn flagged in the company system on the basis of a longer time period that is based on number of top-ups in specific months. The longer time leads to lower accuracy of potential churn customer since they were already gone. Because of that, decision tree models cannot help well for detecting the link between usages and churn rate properly in such circumstances.



TELEKOM SEKTÖRÜNDE FATURASIZ ABONELER İÇİN ABONE HAREKETİ TAHMİNLEMESİ

ÖZET

Telekom sektörü yıllar içerisindeki dijital gelişmelerle birlikte tüm dünyadaki en önemli ve en rekabetçi pazar haline geldi. Telegraf ile başlayan telekomünikasyon sektörü zaman içinde günümüz dünyasında kullandığımız mobil servisler haline dönüştü. Önümüzdeki yıllarda da daha dijital deneyimler ile müşterilerle buluşan, daha da rekabetçi şirketler ile karşılaşacağız gibi gözüküyor.

Yakın geçmişe baktığımızda telekomünikasyonun 4 evreden geçtiğini görüyoruz. Bunlardan ilki, sabit hat olarak adlandırılan kablolu hat ile iletişim sağlanan dönemdir. İkinci evrede ise artık sektör, yavaş yavaş mobil telefona evrilmiştir. Bu dönem şirketlere en çok gelir getiren evredir ve hala sürmektedir. Üçüncü evre ise veri iletişimi/ internet olarak adlandırılan şirketlere en çok gelir getiren diğer bir evredir. Bu devir de günümüzde hala sürmektedir. Veri trafiği arttıkça da sürmeye devam edecektir. Dördüncü ve son evre ise mobil veri evresidir. Telekom pazarının 2005'ten bu yana akıllı telefonların hayatımıza girmesiyle çok yoğun bir şekilde büyüdüğünü görmekteyiz. Bu noktada şirketler, ürün portföylerini genişletmiş, artan talebi karşılamak için fiyat stratejileri belirlemiş ve paralelde ise servis sağlayıcısı için sürekli olarak yatırımlar yapmaya devam etmiştir. Dünyanın dijitalleşmesi ile birlikte mobil telefonlar da artık insanlığın en önemli iletişim aracı olmuştur. Bu noktada, daha da dijitalleşerek her an hayatın içine "bağlı" olmak insanlar için önemli bir hedef olmaya başlamıştır.

Tüm dünyada olduğu gibi Türkiye'de de mobil pazar büyük bir gelişme göstermiş ve müşterilerin sektöre ve dijitalleşmeye bakış açıları değişmiştir. Müşteriler, eskiden ürünlerin fiyatlarına göre seçim yaparken, günümüzde bu durum mükemmel müşteri deneyimine evrildi. Bununla birlikte artık müşteriler ucuz ürün yerine kaliteli ürün talep etmeye başladılar. Bu noktada, şirketler için sadık müşteriler, yeni müşteri edinmekten daha değerli hale geldi. Araştırmalarda ise, yeni müşteri edinmenin, halihazırda olan müşteriyi elde tutmaktan şirketler için çok daha fazla maliyetli olduğu belirtilmektedir.

Telekomünikasyon sektörü yukarıda da bahsedildiği gibi şirketlerin dijitalleşmesinde ve buna ek olarak ülkelerin sosyo-ekonomik gelişmelerinde büyük önem teşkil etmektedir. Bununla birlikte sektör hızla büyümektedir. Türkiye pazarını ele alırsak, sabit müşteriler 11.3 milyon iken, mobil pazar müşterileri 77.8 milyona ulaşmıştır. Bu müşterilerin kullanım detaylarına bakacak olursak, Avrupa ülkeleri arasında en çok konuşan ülkenin 442 dakika ile Türkiye olduğunu görmekteyiz. Türkiye pazarındaki telekomünikasyon şirketlerinin geçtiğimiz yılki pazar payları ise Turkcell 40,5%, Vodafone 37,9% ve Turk Telekom 21,6% olarak raporlanmıştır.

Durum böyleyken şirketler, rekabetçi pazardaki yerlerini sağlamlaştırmak için çeşitli müşteri tutundurma çalışmaları gerçekleştirmekteler. Bu çalışmaların içerisinde en önemlisi de, var olan müşteriyi rakibe kaybetmemek için yapılan çalışmalardır.

Özellikle, sektörde Mobil Numara Taşıma işleminin yaygınlaşması ve kolaylaşmasıyla birlikte, müşteri veri analizi şirketler için daha da önemli bir noktaya taşınmış durumda. Örnekle açıklayacak olursak, 2017 yılı içerisinde 3.3 milyon numara şebekeler arasında taşınmış, bu sayı Mobil Numara Taşıma özelliği hayatımıza girdiğinden bu yana ise 112.3 milyonu aşmıştır.

Müşteri kaybı, telekom dünyasında 3 farklı faza ayrılır. Birinci faz, müşterinin hala şirket müşterisi olarak aktif şekilde kullanım yaptığı evredir. Bu evrede şirket, müşteriden gelir kazanmaya devam eder. İkinci fazda, müşteri hala şirketin müşterisidir fakat kullanım sıklığı azalmıştır. Ek olarak da şirkete kazandırdığı gelir azalmıştır. Bu noktada kullanımların analizleri yapılarak ne zaman gideceği tahmin edilmeli ve ona uygun teklifler sunulmalıdır. Üçüncü ve son aşamada ise, müşteri tamamen bulunduğu operatörü kullanmayı bırakmış ve başka bir operatöre geçmiştir. Bu evreden önce müşterinin gitmesine müdahale edilmelidir.

Müşterinin geçmiş verilerine dayanarak yapılan gelecek tahminlemeleri optimal olarak hesaplandığında, şirketlerin gelir kaybını önlediği bilinmektedir. Bu tahminlemelere göre yapılan müşteri tutundurma aktiviteleri, yeni müşteri kazanımı için harcanan maliyetlerden çok daha verimli olduğundan da literatürde bahsedilmiştir. Ancak, müşteri tutundurma aktivitelerinin de doğru müşterilere yapılması, bir maliyet kalemini daha düşürerek optimizasyonu sağlamamıza yardımcı olur. Bu noktada abone hareketi tahminleme yöntemleri oldukça önem kazanmaktadır.

Abone hareketi tahminleme yöntemleri, şirketlere müşterilerin geçmiş davranışsal verilerini inceleyerek onların diğer operatörlere gidişlerini tahminlemeye yardımcı olur. Bu tahminleme ile birlikte gitme ihtimali yüksek olan müşterileri belirlenir ve bu müşterilere özel yapılacak olan tutundurma aktiviteleri ve stratejileri oluşturulur.

Yukarıda da bahsedildiği gibi, abone hareketi tahminleme yöntemleri şirketlere, rakipler karşısında oldukça avantaj sağlar. Doğru yöntemler ile müşteriye yaklaşıldığında ve kullanımlar analiz edildiğinde müşteriye uygun teklifler sunularak müşterinin rakibe gitmesi önlenir. Bu noktada, telekom sektörü halihazırda gelişmiş bir sektör olduğu ve yeni müşteri kazanımının olan müşteriyi tutmaktan şirkete daha çok maliyet getirdiğini göz önünde bulundurursak, müşterilerden daha fazla gelir sağlayabilmek için şirketlerin var olan müşteri tutundurma aktivitelerine yönelmesi gerektiğini görüyoruz. En temel müşteri tutundurma aktiviteleri genellikle üst tarifeyi önerme, alt tarifeyi önermektir. Bu noktada görüyoruz ki telekom şirketleri aslında müşterinin davranışlarını pazarlama aktiviteleriyle etkileyerek onların daha fazla kullanım yapmasını sağlamaya odaklanmış durumdadır.

Telekom endüstrisinde, diğer bir alternatif pazarlama aktivitesi ise müşteriyi daha sadık tutmak adına faturasızdan faturalıya geçirmektir. Bu stratejideki en önemli nokta, faturalı müşterilerin şirketlere çok daha fazla kişi başı ortalama gelir getirmesidir.

Abone hareketi tahminlemenin doğruluğu şirketler için oldukça önemli bir gelir kalemidir. Abone hareketi hesaplama için birçok program ve method kullanılmaktadır. Genellikle bu metodlar, gıda ya da telekom gibi müşteri sadakatının çok önemli olduğu endüstrilerde kullanılmaktadır. Örnekle açıklayacak olursak, müşterinin dikkatini çekmek ve onu kazanmak sektördeki yarıştan ötürü gıda endüstrisinde oldukça kritiktir. Bu yoğun yarışın sonucu olarak, abone kaybı diğer sektörlerden çok daha fazladır. Bunu önlemek amacıyla şirketler, müşterilerin davranışlarını anlamak ve yeni stratejiler yaratmak zorunda kalmışlardır. Bu noktada da yukarıda bahsedilen abone hareketi hesaplama metodları oldukça önem kazanmış ve sektörde adeta

vazgeçilemez olmuştur. Bu yöntemleri kullanarak şirketler, farklı stratejiler kurup yeni teklifler sunarak müşterilerinin daha sadık olmasını sağlamaya çalışmaktadırlar.

Abone hareketi tahminlemesi için literatürde birçok yöntemden bahsedilmiştir. Ancak yöntemlerden önce abone hareketi tahminlemesi için kullanılacak değişkenlerin belirlenmesi oldukça önemlidir. Bu değişkenlere literatür taranarak ve müşterilerin kullanımları incelenerek karar verilebilir. Literatürde en önemli abone hareketi tahminlemesi değişkenleri; indirim oranı, servis kalitesi, arama kalitesi, tarife ücreti, müşteri desteği olarak bulunmuştur. Ancak bunlara ek olarak, müşterilerin veri, ses, dakika kullanımları, yurt dışı arama oranları, diğer operatörleri arama sıklıkları, kontör yükleme sıklığının da önemli değişkenler olduğundan bahsedilmiştir.

Bunlara ek olarak, çeşitli araştırmalarda, dakika, veri ve SMS kullanım oranları, aylık fatura ödeme tutarı ve faturasız müşteriler için aylık kontör yükleme sıklığı abone hareketinin tahminlemesinde birbirleriyle ilişkisi olan değişkenler olduğundan bahsedilmiştir.

Kullanımlar ve ödeme planı haricinde davranışsal sadakat ve müşterinin operatörde bulunma süresi, şirkete olan memnuniyeti pozitif etkilemektedir. Ancak, her ne kadar durum böyle olsa da, ürün ve servislerden memnun olan müşteriler kesinlikle operatöre sadıktır gibi bir yorum yapamayız. Müşteri memnuniyeti ve operatörde bulunma süresi doğrusal ve basit bir ilişki değildir. Çünkü operatörde bulunma süresi üzerinde ürün/servis kalitesi, fiyat algısı, sosyal çevre ve kontrat süresi gibi diğer etkenler de vardır. Buna ek olarak, bağlılık ve memnuniyet arasındaki ilişki müşteri memnuniyet skoruna da bağlıdır. Örneğin, düşük memnuniyet skoru veren bir müşteri, başka bir operatöre gitmeye yatkındır gibi yorumlanabilir.

Tezin içeriğini oluşturan faturasız müşteriler genellikle, göçmenler, genç segment ve harcamalarını kontrollü yapmak isteyen segmentler tarafından tercih edilmektedir. Faturasız hat kullanımı, mobil numara taşımanın kolaylaşmasıyla daha da yaygınlaşmıştır.

Bu tez içerisinde faturasız müşterilerin geçmiş verileri SPSS V22.0 programı üzerinde Karar Ağaçları yöntemiyle incelenmiş ve sonuçlarından bahsedilmiştir. Sonuçta, davranışsal ve demografik verilerin abone hareketi tahminlemesini etkilediği ortaya konmuştur. Özellikle, diğer operatörlerdeki müşteriler ile konuşma süresi, tarife aşım süresi, müşterinin şirket bazında bulunma süresi ve yaşı seçilen değişkenler içerisinde en yüksek doğruluğu veren değişkenler olarak bulunmuştur.

Karar Ağaçları yöntemini veriye uygulamadan önce, değişkenlerin birbirleriyle olan ilişkileri incelenmiş ve veri üzerindeki dağılımlarına bakılmıştır. Bu dağılımlara göre değişkenler gruplandırılarak yorumlama kolaylaştırılmıştır. Birbirleri ile ilişkili çıkan değişkenler, Karar Ağaçları yöntemi sonucunda çıkan verilerle de karşılaştırılmıştır.

Bunlara ek olarak, faturasız müşterilerin müşteri bazından ne zaman tamamen gittiğinin bilinmemesi ve şirket sistemlerinde müşterilerin “giden” olarak işaretlenmesinin belirli bir süre kontör yüklememesi durumuna göre verilmesi, davranışsal verilerin doğruluğunu olumsuz etkilemektedir. Bu sebeple giden müşterilerin kullanım bazlı hareketleri Karar Ağacı'nı etkilemiştir.



1. INTRODUCTION

Telecommunication sector plays an important role for the digitalization of the companies and socioeconomic development of the nations in the last decade. Therefore, mobile market has a remarkable growth in Turkey and customer perception being changed. Customers used to act price sensitive in recent years. But now, they are demanding perfect product experience and quality as well. Under these circumstances, creating loyal customers are much more important than acquiring new customers. According to researchers, acquiring new customers overpriced compared to keeping existing ones (Altas and Gülpinar, 2013).

According to published studies, the usage behaviors have been represented by measurement of MoU (minutes of usage), the rate of usage and total number of incoming calls, (Wei and Chiu, 2002), monthly usage (Buckinx and Vanden Poel, 2005) and Mozer et al. (2000) study shows that churn rate is correlated with usage rate and the monthly charges.

Same operator incoming or outgoing calls cost less compared with other operators' calls. It can show that if a customer make more call with other operators' subscribers, (s)he tends to choose the operator which make more call. It is understood that social network (relationships) has a very significant role in churn management (Dasgupta et al., 2008). In addition, if the rate of VAS Usage Revenue to Total Usage Revenue is greater than 30 percent, it is assumed that consumers are disposed to churn (Geetha and Abitha Kumari, 2012).

According to Noyan and Simsek (2014), important components of the customers' loyalty are discount rate and service quality. Also, as stated in Zakaria et al. (2014), customer loyalty increase with reward and discount rate. For example, if the customer gets reward for purchasing specific number of goods, customers tend to buy more and reward plays a significant role for being less cherner.

According to Bolton (1998) and Gerpott et al. (2001), satisfaction score of mobile carriers' service has three main components: call quality, tariff price and customer

support. For instance, according to Gale (1992); %90 of AT&T customers rated the service quality as great and it represents the repurchase ratio, while other customers rating as good, fair or poor (60%, 17% and 0%, respectively) have low repurchase disposition. Besides, service quality experience is affected by some changes (e.g., from fair to good). This experience generally changes the decision of repurchase exceedingly in comparison with other factors. In this point, it should not be forgotten that because of subjectivity of the satisfaction level, it cannot be classified equally by customers. For instance; while call quality is a key factor for churn behavior both Germans and Koreans, price and customer support is important only for either Germans or Koreans. (Gerpott et al., 2001; Kim et al., 2004; Kim & Yoon, 2004).

Based on statistics, total revenue of total telecommunication sector increased from 147 billion USD in 1980 to 1200 billion USD and it is expected that it will be reach to 1290 billion USD at the end of 2019. This accelerated growth is directed mobile operators to innovation in the services and try to find some ways to keep the customers in their company. In a way that this supports telecommunication sector is becoming more competitive with tariff prices, regulations and MNP (Mobile Number Portability). Mainly brand switching trend in telecommunication sector push the companies to find revenue-increasing activities.

CAPEX trend is increased at the same rate with global industry revenues (%5 CAGR (Compound Annual Growth Rate) in 2005 including spectrum costs for investments. Because of increased smartphone penetration, mobile sector has become the lead of the digital sector indubitably.

The focus of this paper is to predict of customer churn using demographic and behavioral customer data using Decision Tree algorithm. In addition, selected determinants of the customer churn were analyzed according to churn status.

The structure of this paper is organized as follows. In Section 2, global and Turkish telecommunication evaluation in years and effects with numbers. In section 3, customer churn definition, types and phases, in section 4, methodology and previous researches about churn prediction methods. Section 5 presents implementation, findings, limitations and finally, in section 6, it is discussed that conclusion of the model and further suggestions for future studies.

2. OVERVIEW TO TELECOMMUNICATION SECTOR

2.1 Evolution of the Telecommunication Market

The telecom sector includes companies which give a global communication service to consumers operates through telephone or Internet. These firms' IT infrastructure provides transferring of data, call services and also short messages globally.

Telegraph advance to mobile services in time in the telecom sector and while communication took days with telegraph, extremely high data amount can be sent quickly at the present time due to technological evolvement.

World's telecom sector evolved to phase of extensive changes in 1980's. Indeed, this phase of extensive changes was established by United States government who has own PTTs (Post, Telegraph & Telecommunication) in the late nineteenth and early in the twentieth century. Along the years, this situation was investigated and the problem only could be solved that communication sector can change without any limit by government monopoly. In this way, the "American Telephone & Telegraph" (AT&T) was split up in 1984. The new era has begun for telecom sector with this division of the AT&T and it provides full competition among the telecommunication companies in USA (Lindmark et al., 2006, pp. 46).

British Telecom that held monopoly position until 1996 in U.K. market was corporatized and faced with much competitions. The liberalization trend in the telecom sector brought many important results. The new wireless companies demanded to use most advanced systems to be able to take competitive advantages from potential customers. Operators, which are not only privatized but also ex monopoly operators, had to improve these systems in terms of technique and economy, in order to decrease customer loss. Quick evolvement of the system innovations, competition between operators and increase of net traffic growth in the market in the world bring about high growing of the telecommunication sector. The companies were expecting that the

aggregate annual growth rate is rise from \$588 billion in 1995 to \$1,06 trillion in 2000 (Ploetner, 2004, pp. 344-345).

A phenomenal change for long-established firms where produced telecommunication equipment has been made. Main suppliers, who are generally local ones, have been preferred by old monopolists regarding technology and procurement processes. On the other hand, operators started to make pressure to their suppliers. These suppliers faced with some challenges effect of liberalization wave: Covering the intensely changing demands of their customers and focusing on much stronger in international business.

As seen Figure 2.1 (Lindmark et al., 2006, pp. 50), dependent on two important trends, which are towards mobile telephony and data communications, four different segments, were defined:

- Main and evolved segment: Fixed Telephony,
- Revenue generator segment: Mobile Telephony,
- Another revenue growth- generating segment (especially data traffic) which has disruptive effect on the market: Data Communications/Internet
- Revenue growth potential: Mobile Data

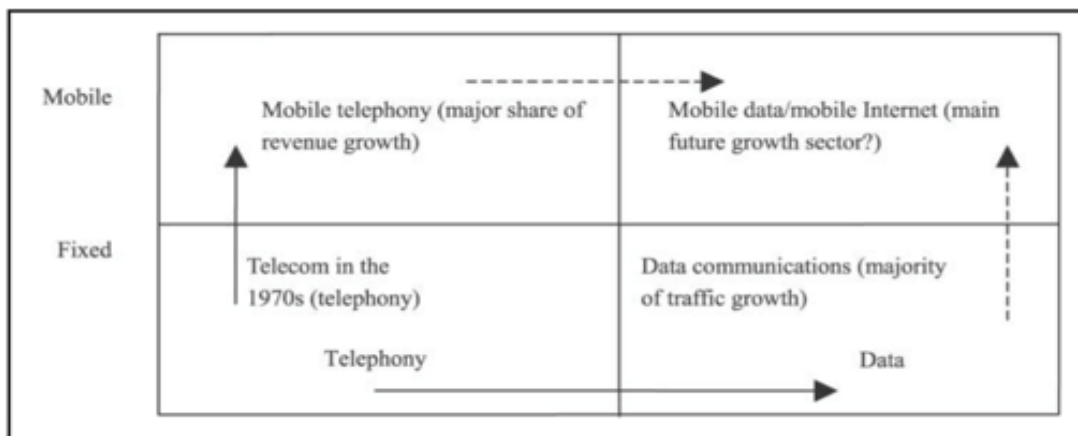


Figure 2.1: Main growth trends in telecom in 1970-2003.

Importance on economic growth of these segments was relatively transform in time and it is expected that this evolution will continue mainly shaped by innovation (Lindmark et al., 2006, pp. 50).

The telecom business evolved extremely since 2005 because of consumers who consumed data enormously with their smart devices. In this time, operators have extended their product portfolios and arranged its prices to cover the excessive demand with increasing network capability expenditures. Operator performance relies on the region. North America and Asia regions have outstanding performance comparing with other regions in terms of using increased scale and powerful secular growth opportunities, respectively. However, Operators in Europe region performance have increased on expectation of the merger, whereas Latin America region carriers have affected the worsening macroeconomic status.

Furthermore, most of the operators are comprehend the revenue channels, it is a challenge whether new product and services provides profit or not. Also, firms divide their business using new offerings, discompose the elements of the industry in process. Stakeholders of the digital environment are trying to find that new ways for differentiation in order to increase the consumer spend of money. During operators still appreciate the ecosystem revenues; OTT contents have increased their revenue share up to 10% in a couple of years, also competition between companies in retail and distribution got more excited. The telecom sector has diversified without being noticed since 2005 together with customer needs and competitive situation are differentiated in many ways, which are not predictable.

EY Global Telecommunications study (2015) shows to findings of the industry and what expecting about future.

2015 key findings of global telecommunication industry:

- Destructive competition is the main challenge for industry.
- OTT is the one of the popular content for changing the customer demand.
- Regulative uncertainties still disconcert to sector.
- Customer Experience is hot topic for the companies strategically.
- Quality of service and customization are one of the key factors for customer focus.
- Network Quality is the distinctive factor of differentiation.
- A new interaction between customers and processes can increase agility level.

- One of the main drivers of the M&A is in market consolidation.
- Digital services will be key revenue-generator until 2020.
- While IoT revenue growth expectation is low, TV and cloud have more certainty.

2.2 Turkey Market Statistics

Telecom sector is the main driving sector for all data related industries in the last few years and all data gathered from mobile customers force other sectors to co-operate with operators. Non-telecom sectors are going to be ambulance chaser behind the telecom sector. As in the example of Yandex, Technology Company models its navigation app through using the customer location data.

According to records, while all market players' Net Sales Revenue was 45.4 billion TL in 2016, Net Sales Revenue was 51.1 billion TL at the end of 2017.

In Turkey market, while number of fixed customers was 11.3 million, mobile services were extended to 77.8 million subscribers and also mobile penetration rate has reached to 96,3%.

In the 3rd quarter of 2017, average MoU (Minutes of Usage) per month was 442 minutes which is the first place among the European countries. At the same time, 3.3 million number has ported during this period. On the other hand total number of Mobile Number Portability (MNP) has exceeded 112.3 million until today.

Broadband Internet subscribers, which were around 6 million in 2008, increased to more than 68,9 million as from end of 2017.

When market shares of the carriers were examined in detail according to 2017 results, market shares of Turkcell, Vodafone and Avea were 40,5%, 37,9%, and 21,6% respectively (Bilgi Teknolojileri ve İletişim Kurumu Pazar Verileri Report 4th Quarter, 2017).

3. CUSTOMER CHURN

Customer churn efficiency is one of the most important variable for generating income for companies. Churn management tools are commonly used in various industries such as food industry or telecommunication sector.

Attracting and gaining the customers in the food industry is notably critical because of the competition in the sector. As a result of this intense competition and entity of several stores, customer churn rate is much more higher than the other sectors (Khodabandehlou and Rahman, 2017). These companies try to understand their customer behavior and create strategies for gaining customers. In this point, churn management methods are becoming irreplaceable for this industry. With these methods, companies can monitor their customers' behavior and they can provide better services to keep them loyal.

Companies in the competitive markets, should identify their retention strategies and decide method of customer churn management. For example, Business Intelligence (BI) tools are used in the businesses for better decision-making (Turban et al., 2008). These tools (e.g. data mining or machine learning) not only help them to categorize the customers as churner or non-churner but also lead to targeted marketing strategies for potential churner customers. Previous studies about churn prediction modelling is also shown in Table 3.1.

Vafeiadis et al. (2015) study about frequently used churn prediction techniques which are ANN, Support Vector Machines, Decision trees, Naïve Bayes and Logistic regression and evaluation of their performances using performance classifiers in telecommunication industry. Comparative results indicated that the best method named boosted SVM for the churn prediction in telecom industry was founded.

Inoue et al. (2017) mainly focused on pattern of customers through corporate social responsibility. Frequency of attendance of the social responsibility was affected by gender, age, tenure and perceived performance.

Kim and Moon (2012) study aims to create successful retention activities for win the customers back. Developed retention models which are based on churn probability, retention probability, expected revenues and combination of these models are compared between them using different algorithms. Model CP which based on churn probability was determined as most accurate model for churn forecasting.

Amin et al. (2017) study aims to forecasting of the customer behavior while using rule generation methods which are Exhaustive Algorithm, Genetic Algorithm, Covering Algorithm and LEM2 Algorithm implemented on benchmark telecom dataset. Rough Set Classification was applied for measurement of the methods. As a result of this process, Genetic Algorithm gave the best churn probability rate.

Zorn, Jarvis and Bellman (2010) study mainly mentioning about different perspective of churn prediction methods while using a DVD firm customer dataset. Cox regression model was implemented for this study and variable effects were investigated in depth. In conclusion, payment per month had a remarkable impact of customer churn rate.

Customer churn explained in literature in Telecom industry that subscription from existing company is ended up by the customer and transportation of number or subscription to another carrier is entitled 'churner'. Since, exponential increase of the competition among telecommunication companies, customer churn become significant problem for the companies which are not only in telecom industry but also insurance and banking.

Three types of customer churn are described in the literature as seen below (Lazarov and Capota, 2007):

- Active Churner (Volunteer): Customers who prefer to quit existing carrier and want to switch another carrier. Termination reasons can be listed, for instance, subscription termination, mobile phone device upgrade/downgrade, network quality, competitors marketing activities, regulations etc.
- Passive Churner (non-volunteer): Customers who are thrown by the company because of the irregularity in invoices.
- Rotational Churner (Silent): The contract is ended up by company or customers unexpectedly without any pre-notification.

Table 3.1: Previous studies about churn

| Author | Source | Factors | Methodology | Results |
|--------------------------------|---|--|--|--|
| Vafeiadis et al.(2015) | Simulation Modelling Practice And Theory | No. of months active user Total charge of evening/day/night/international calls Total minutes of evening/day/night/international calls Total number of evening/day/night/international calls No. of voice-mail messages/customer service | ANN Support Vector Machine Decision Trees Naive Bayes Logistic Regression | Comparison between the algorithms are shown. Accuracy of ANN, Support Vector Machine and Decision Tree is higher than Niave Bayes and Logistic Regression. |
| Inoue et al. (2017) | Journal of Business Research | Gender Age Commitment Involvement | Confirmatory Factor Analysis | Attendance to the Corporate Social Responsibility can only moderate by objective data. CSR initiatives can increase behavioral loyalty. |
| Kim and Moon (2012) | Expert Systems with Applications | Churn probability Retention probability Churn and retention probability Successful retention satisfaction Switching cost | Naive Bayes Neural Networks Decision Tress Random Forest Logistic Regression | Cultural differences demographic variables and social characteristic of the customer can impact of the satisfaction and switching costs. |
| Amin et al. (2017) | Neurocomputing | Existance of international plan Existance of Voice mail plan Customer service calls No. of minutes international calls | Genetic Algorithm Exhaustive Algorithm Covering Algorithm LEM2 Algorithm | Genetic algorithm with RST classification results are much more accurate compare with other methods. These results can be based on retention activities. |
| Zom, Jarvis and Bellman (2010) | Journal of Research in Interactive | No. of customer visit to the website No. of up or downgrades between offers Average rental time Number of title reviews Payment per month | Cox-regression Analysis | Visit count, number of title reviews and average rental time does not have significant effect on behavioral data. But, payment per month has. |

Customer churn types except rotational churning are predicted using several approaches in the literature. Rotational churning is not easily predictable, because it depends on probability.

In Figure 3.1 (Geetha and Abitha Kumari, 2012), as seen below, the customer churn modeling and also shaded region specified in this study. Majority of past researches targeted on to determine the direct effect of independent variables on customer churn except Ahn et al. (2006) who examine the consumer position as an intermediary for churn determinants and customer churn. Consequently, any customer churn could occur in three stages, which are shown in Figure 3.1. The initial and final stages are explained in details in the literature. Nevertheless, the intermediary stage consists of active and non-active states. For instance, the company can not bring in revenue in dormant stage from the customers who do not use network. (Ahn et al., 2006).

Other phase is examined as seen in Figure 3.1. This is called as the revenue churn phase or dormant phase. The stream of consumer from revenue earning to non-revenue earning can be defined as revenue churn. Non-revenue earning consumers (NREC) are the class of consumers in this related phase who do not recharge, makes calls or do not operate any of the services of the service provider. This class of consumers are categorized as present but dormant. These consumers are intended to generate revenue loss to the firm. Thus, preventing revenue loss of these, NREC and to evaluate their delicacy to churn, the usage routine has to be examined.

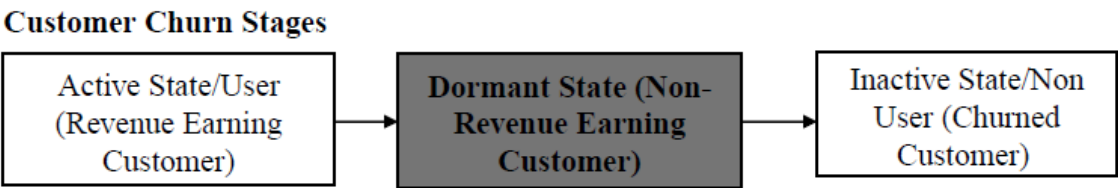


Figure 3.1: Phases of churn.,

When looked up to overall performance of the market, SAS Institute (2001) published that the cost of churn is approximately \$4 billion per year in EU and US telecommunication market, also global costs reached \$10 billion each year.

It has been observed that company’s performance is affected by customer churn as described below (Amin et al., 2017):

- Negative impact on company performance in terms of customer number.
- Another impact on sales is that new customers are willing to purchase less services compared to others.
- Business promotions by competitors' effect unsatisfied customers' decisions on carrier services.
- Customer churn has a negative impact on companies' revenues.
- It might alter with long-term customer's opinion.
- It can cause unpredictability, which decreases the ratio of the potential new customers.
- Gaining new customers from other carriers are more costly when compared to preserving existing customers.
- Company's reputation is effected from losing customers to competitors.

There is a direct connection between customer loss and loyalty. In today's business world, price cuts are not enough for loyalty. Customers become more loyal if the companies add a new value to the original product. The main objective for churn analysis is to identify the customers likely to be lost and compute the total cost for the effort to win them back.

3.1 Behavioral Loyalty

Churn rates can be 30% per year in saturated markets. (Thomas et al., 2004; Wei and Chiu, 2002). As explained above sections, churn rate defined as ratio of disconnected customers from their company. Because of the high churn rates, companies should create loyalty to their products (Kamakura et al., 2003). On the other hand, higher churn rate causes the decreasing of customer acquisition (Bonfrer et al., 2008). However, as it mentioned before, customer acquisition activities has high expenditure and acquiring each customer back is not profitable. In some cases, acquisition activity costs can be higher than the customer lifetime ARPU (Average Revenue per User) especially for temporary customers (Reinartz and Kumar, 2002; Novak and Hoffman, 2000). Owing to long commitment period and product awareness, long-term customers can be more profitable than the short-term customers (Reichheld, 1996). In addition,

customer acquisition number is affected by negative word of mouth from detractors of the company (Allenby et al., 1999; Lopez et al., 2006).

Brand loyalty is defined in previous studies as loyalty of customers to a specific brand without any dependency to marketing action of the company (Oliver, 1999). It has two types loyalty as attitudinal and behavioral. While, attitudinal loyalty is described as consumers' attitude against particular brand, behavioral loyalty is defined as repurchase tendency for a particular products or services (Kumar and Reinartz, 2006). It has been also described that loyal customers in terms of behavior are not always act attitudinally loyal and competitors cannot faced with any difficulty while enticed to customers for purchasing their products (Giovanis and Athanasopoulou, 2018) Also, behaviorally loyal customers cannot be genuine customers in repurchasing activity of the products or services just due to their specific situations.

Companies have challenges about accessing behavioral data. Only few data can easily accessible, such as call frequency in telecom industry (Wei and Chiu, 2002). Although, churn rate can be calculated adding other affective measurements such as customer characteristics and attitude (e.g satisfaction score obtained via surveys). However, satisfaction score is occasionally added as measurement to the data considering that it is sample of total customers and not real time. On the contrary, behavioral data is consistently up-to-date for each customer (Zorn et al., 2010).

Prediction of customer behavior and customer value is made more accurately with past behavioral data in comparison with customer demographics (Hughes, 2005). Hence, score models on historical purchasing data (e.g customer billing) is used for calculation of customer value and classify the customers accordingly (Reinartz and Kumar, 2002).

Some researchers consider that behavioral loyalty and customer tenure affect to satisfaction score positively (Dover and Merthi, 2006; Ganesh et al., 2000). Nevertheless, customers, who satisfied with the product or services, have not certainly loyalty to company (Oliver, 1999). Generally, satisfaction and customer loyalty does not have a linear and simple relation because of other effects on customer tenure, such as product/services quality, price perception, social network (Oliver, 1999) and commitment period (Bolton et al., 2004). In addition to this, relationship between satisfaction and loyalty is affected by customers' satisfaction judgement. For example, low satisfaction score is less likely to transform to customer loyalty (Chandrashek et

al., 2007). Another example is that if the companies create service failure uncertainty, satisfaction score of the company decreases, even after the problem is solved (Zeithaml et al., 1996).

3.2 Churn in Telecom Industry

New markets are developing in economy, as a consequence of this competition evolving between companies. Evolving competition, force companies to make more innovation. For survival, they need to differentiate. Customer care is the ultimate answer for loyalty and customer satisfaction.

Predicting the consumer behavior provides a core competence for the companies. Analyzing and implementing a related model for customer needs is essential to maintain the relationship between the consumer and the company. Since, the telecommunication sector is already a developed market and gathering customers from competitors are overpriced, companies are more focused on existing customers, which are named cross-selling and up-selling in Customer Relationship Management in order to gain more profit. Therefore, telecommunication companies try to affect customer's behavior with marketing activities for managing the usage of tariffs.

Commonly, prepaid tariffs are preferred generally by immigrants, youths and customers who would like to control their spending. This trend is becoming popular due to easy number carrying options while switching the brands.

According to Kisioglu and Topcu (2011) study results, for instance, for an efficient churn management in telecom industry, average MoU (Minutes of Usage), average invoice payment, number of interconnection calls and tariff type are most significant factors in order to analyze customer churn rate.

3.3 Churn Management Methods

Churn management is operating in several sectors with dynamic actions. Each sector has different methods for dealing with churn customers. Churn management should be considered analytically. Consumer behaviors and actions leave a hint for the companies as they can be tangible to churn in upcoming period. Sectors including from online broadcasting to online retails, also from credit card services to carrier providers and etc. have different symptoms for customer churn prediction and the

different solutions should be implemented for each sector. In online broadcasting business such as Netflix, customers are mostly contractual customers, and they paid the service with their credit cards, based on consumer reports consumers have not paid their last two subscription fees are tangible to cancel the service. Netflix detect these consumer and offer a reasonable deal for future service. In non-contractual sectors, such as online fashion retailers (e.g. Alibaba), consumers are tangible to churn if they do not have make any action at the platform for two months time. To avoiding churn in this manner, online retailers offer them a deal with their possession of interest. In the type of voluntary churn, mostly credit cards are facing with this consume churns and it is hardly detect before the action. However, for voluntary churns, credit card companies, create a method of “rewarding”. When a consumer wants to leave, these companies offer a last time deal such as bonus credit balance for shopping. Involuntary churn is another problem mostly credit companies faced with. When a consumers credit card expires, it involuntarily dropped out from the system, thats why credit card companies send new credit cards months before expiration dates.

In telecom industry, one of the alternative activity for turn it to advantages is pre-to-post strategy which means activities for prepaid customers to migrate to postpaid, in order to gather customers that are more loyal. Most important point in this strategy is postpaid customers bring much greater ARPU (Average Revenue per User) than the prepaid customers. Another alternative activity is that focusing on loyalty of existing customers. Many studies are written up both behavioral an attitudinal loyalty of prepaid and postpaid customers; also evaluate the customers using data mining in the literature.

When looked up the Turkish Telecom sector, it is found that important relationship between subscription type and customer income level (Aydin, 2004). Another survey on the graduate students indicated that important relationship is found between subscription type and objects like faculty, graduation year, age, gender, education level of the family members and income level of the family (Oktay et al., 2006). Also, other researches are indicated that prepaid customers are more effected from price changes when it’s compared to postpaid customers (Karacuka et al., 2011). Other results of the researches, behavior of the peers about the brand switching also shows that social network of the customer make much more remarkable impact on subscribers’ brand decision when it is compared to mobile network quality.

4. RESEARCH DESIGN

4.1 Aim and Scope of the Research

As mentioned in previous section, carriers can use several retention activities to prevent customer churn. However, it is the key point for when and how will be touched to customers. Especially behavior of prepaid customers who have low commitment rate hard to examine.

This paper aims to predict to churn rate of prepaid customers, and create an action plan before they abounded to carriers. An algorithm is selected and implemented for analysis. A dataset is obtained from a Turkish telecom company for determined period and results are examined as seen below sections.

4.2 Data Preparation and Sampling

In first step, a dataset is obtained from a carrier including churning and non-churning customers' behavioral data, which has totally 429 samples. In Table 4.1, type of customer and sample numbers are seen below. Churning customers are selected from March 2018 data randomly, which was totally 400 customers sample. However, for gathering more accurate results, behavioral data is collected from December 2017 and 127 customers have movement on December'17. Prepaid customers churn flag is given 9 months after last top-up. Therefore, remaining 273 churning customers were already gone from carrier customer base before getting churn flag.

Non-churning customers are collected from another data, which is called "Net Promoter Score" time between April'17 and November'17. Detractors of the company who are willing to abandon to company is selected and 302 non-churning customers' Dec.'17 behavioral data is examined as seen Table 4.1.

Table 4.1: Training dataset with numbers

| Customer Type | Sample Number | Data Time Period |
|---------------|---------------|------------------|
| Churner | 127 | Dec.'17 |
| Non-churner | 302 | Dec.'17 |

After selection of customer sample and period, variables of the behavioral data has been chosen based on recent researches on literature, which mentioned above sections and 15 different variables has been examined as seen below Table 4.2. Defined variables are effected churn and contribute to prediction of churn rate.

Table 4.2: Definition of the selected variables

| Variables | Data Type | Defination |
|---------------------------------|-------------|--|
| Age | Demographic | Age belong to user |
| Tenure | Demographic | Months stay on the carriers' base |
| Duration Usage Rate | Behavioral | Total minutes of usage divided by tariff bundle (giving from carrier within the scope of tariff package) minutes |
| Data Usage Rate | Behavioral | Total data usage divided by tariff bundle (giving from carrier within the scope of tariff package) data |
| SMS Usage Rate | Behavioral | Total SMS usage divided by tariff bundle (giving from carrier within the scope of tariff package) SMS |
| OOB Duration Rate | Behavioral | Total minutes of usage removal from tariff bundle (giving from carrier within the scope of tariff package) divided by total minutes of usage |
| OOB Data Usage Rate | Behavioral | Total data usage removal from tariff bundle (giving from carrier within the scope of tariff package) divided by data usage |
| OOB SMS Usage Rate | Behavioral | Total SMS usage removal from tariff bundle (giving from carrier within the scope of tariff package) divided by SMS usage |
| Offnet Usage Rate | Behavioral | Other carriers customers' call divided by total outgoing call |
| Incoming/Outgoing Duration Rate | Behavioral | Call record interaction rate between recipient and caller |
| VAS Product Rate | Behavioral | Number of value added services (VAS) belong to customer divided by total VAS |
| # of Migration (annual) | Behavioral | Number of tariff change in a year |
| # of Top-up (annual) | Behavioral | Number of added airtime minutes to customer line in a year |
| # of Port in (annual) | Behavioral | Number of subscribed times to selected carriers' base in a year |
| # of Port out (annual) | Behavioral | Number of unsubscribed times to selected carriers' base in a year |

4.2.1 Dataset interpretation

Variable similarities and differences on churner and non-churner data are analyzed separately. As seen in the Table 4.3, customer behavior in terms of offnet usage, almost equal to onnet usage on non-churner data sample. On the other hand, offnet usage of the customers is higher than onnet usage on churner data. This variable may have encouraged the customer to abandon the carrier.

Table 4.3: Offnet usage rate similarity

| Offnet Usage Rate Similarity on Non-churner Data | | Offnet Usage Rate Similarity on Churner Data | |
|--|--------------|--|--------------|
| Offnet Usage Rate | # of samples | Offnet Usage Rate | # of samples |
| 0 | 34 | 0 | 53 |
| 0-0,5 | 107 | 0-0,5 | 33 |
| 0,5-1 | 161 | 0,5-1 | 41 |

As seen in Table 4.4, OOB usage is lower than total minutes usage not only in non-churner, but also in churner data. In addition, customers generally keep away from exceeding their tariff as expected.

Table 4.4: OOB usage minutes similarity

| OOB Usage minutes Similarity on Non-churner Data | | OOB Usage minutes Similarity on Churner Data | |
|--|--------------|--|--------------|
| OOB Usage minutes | # of samples | OOB Usage minutes | # of samples |
| 0 | 149 | 0 | 53 |
| 0,01-0,1 | 58 | 0,01-0,1 | 18 |
| 0,1-0,5 | 34 | 0,1-0,5 | 15 |
| 0,5-1,00 | 61 | 0,5-1,00 | 41 |

Another analysis about customer behavior as seen below, while customer tenure is higher on non-churner data, customers are willing to abandon the company in a year on churner data.

Table 4.5: Tenure similarity

| Tenure (month) Similarity on Non-churner Data | | Tenure (month) Similarity on Churner Data | |
|---|--------------|---|--------------|
| Tenure | # of samples | Tenure | # of samples |
| 0-12 | 52 | 0-12 | 68 |
| 13-27 | 105 | 13-27 | 35 |
| 28+ | 145 | 28+ | 24 |

Youth and older segments have shown same behavior on non-churner data and they willing to stay in their existing carrier. However, older segment willing to abandon the existing company on churner data.

Table 4.6: Age similarity

| Age Similarity on Non-churner Data | | Age Similarity on Churner Data | |
|------------------------------------|--------------|--------------------------------|--------------|
| Age | # of samples | Age | # of samples |
| 0-26 | 123 | 0-26 | 15 |
| 27-35 | 51 | 27-35 | 35 |
| 36+ | 128 | 36+ | 77 |

Non-churner customers made top-up nearly twice a month while churner customers did not make top-up even every month. Also, 15 customers in churner data has made top-up twice a year.

Table 4.7: Top-up similarity

| Top-up Similarity on Non-churner Data | | Top-up Similarity on Churner Data | |
|---------------------------------------|--------------|-----------------------------------|--------------|
| # of Top-up (annual) | # of samples | # of Top-up (annual) | # of samples |
| 0-15 | 268 | 0-15 | 93 |
| 16-24 | 32 | 16-24 | 34 |
| 25+ | 2 | 25+ | 0 |

As seen in Table 4.8, rate of incoming/ outgoing calls are similar on churner and non-churner data. It is proven that, outgoing calls is higher than incoming calls in the dataset.

Table 4.8: Rate of incoming/ outgoing calls similarity

| Rate of incoming/outgoing Calls Similarity on Non-churner Data | | Rate of incoming/outgoing Calls Similarity on Churner Data | |
|--|--------------|--|--------------|
| Incoming/Outgoing | # of samples | Incoming/Outgoing | # of samples |
| 0-0,5 | 51 | 0-0,5 | 22 |
| 0,5+ | 251 | 0,5+ | 105 |

Minute usage and SMS usage rate distribution are similar in the data. Most of costumers in churner and non-churner data are using their tariff content with the percentage of % 50 as seen in the Table 4.9.

Table 4.9: Duration and SMS usage rate similarity

| Duration Usage Rate Similarity on Non-churner Data | | Duration Usage Rate Similarity on Churner Data | |
|--|--------------|--|--------------|
| Duration Usage Rate | # of samples | Duration Usage Rate | # of samples |
| 0-0,5 | 150 | 0-0,5 | 62 |
| 0,5-1,00 | 96 | 0,5-1,00 | 41 |
| 1,00 + | 56 | 1,00 + | 24 |

| SMS Usage Rate Similarity on Non-churner Data | | SMS Usage Rate Similarity on Churner Data | |
|---|--------------|---|--------------|
| SMS Usage Rate | # of samples | SMS Usage Rate | # of samples |
| 0-0,5 | 259 | 0-0,5 | 109 |
| 0,5-1,00 | 10 | 0,5-1,00 | 4 |
| 1,00 + | 33 | 1,00 + | 14 |

While data content of the tariffs are used more than percentage of %50 in non-churner data, churner customer usage distribution is irregular as seen Table 4.10.

Table 4.10: Data usage rate similarity

| Data Usage Rate Similarity on Non-churner Data | | Data Usage Rate Similarity on Churner Data | |
|--|--------------|--|--------------|
| Data Usage Rate | # of samples | Data Usage Rate | # of samples |
| 0-0,5 | 32 | 0-0,5 | 53 |
| 0,5-1,00 | 108 | 0,5-1,00 | 33 |
| 1,00 + | 162 | 1,00 + | 41 |

Non-churner customers try to find a suitable tariff for themselves and staying in the existing carrier, while churner customers can either abandon the company or not as seen in the Table 4.11.

Table 4.11: # of Migration similarity

| # of Migration Similarity on Non-churner Data | | # of Migration Similarity on Churner Data | |
|---|--------------|---|--------------|
| # of Migration | # of samples | # of Migration | # of samples |
| 0 | 110 | 0 | 69 |
| 1+ | 192 | 1+ | 58 |

Non-churner customers do not move to other carrier mostly in selected year or come from another carrier. They can mostly labelled as loyal customers. Churner customers in the dataset act similar with non-churner data unexpectedly as seen Table 4.12.

Table 4.12: # of Port-in and Port-out similarity

| # of Port-in Similarity on Non-churner Data | | # of Port-in Similarity on Churner Data | |
|--|---------------------|--|---------------------|
| # of Port-in | # of samples | # of Port-in | # of samples |
| 0 | 253 | 0 | 91 |
| 1+ | 49 | 1+ | 36 |

| # of Port-out Similarity on Non-churner Data | | # of Port-out Similarity on Churner Data | |
|---|---------------------|---|---------------------|
| # of Port-out | # of samples | # of Port-out | # of samples |
| 0 | 294 | 0 | 122 |
| 1+ | 8 | 1+ | 5 |

As seen in the Table 4.13, VAS Product Usage Rate is not cumulative on 0 or 1 neither in non-churner nor churner data.

Table 4.13: VAS product rate similarity

| VAS Product Rate Similarity on Non-churner Data | | VAS Product Rate Similarity on Churner Data | |
|--|---------------------|--|---------------------|
| VAS Product Rate | # of samples | VAS Product Rate | # of samples |
| 0 | 145 | 0 | 71 |
| 0-1,00 | 157 | 0-1,00 | 56 |

Most of customers in not only non-churner but also churner data are not exceeding their tariffs excessively in terms of Data and SMS content.

Table 4.14: OOB Data and SMS usage similarity

| OOB Data Usage Similarity on Non-churner Data | | OOB Data Usage Similarity on Churner Data | |
|--|---------------------|--|---------------------|
| OOB Data Usage | # of samples | OOB Data Usage | # of samples |
| 0-0,5 | 229 | 0-0,5 | 104 |
| 0,5-1,00 | 73 | 0,5-1,00 | 23 |

| OOB SMS Usage Similarity on Non-churner Data | | OOB SMS Usage Similarity on Churner Data | |
|---|---------------------|---|---------------------|
| OOB SMS Usage | # of samples | OOB SMS Usage | # of samples |
| 0-0,5 | 292 | 0-0,5 | 108 |
| 0,5-1,00 | 10 | 0,5-1,00 | 19 |

4.2.2 Relation of variables

Dependency to the churn status of each variable has been analyzed using SPSS v22. Churn status (churner or non-churner) has been accepted as a constant and Hypothesis was defined as seen below. Each variable has been tested based on this constant using Pearson Chi-Square and Mc Nemar as seen below tables.

Hypothesis (H0): Variables and churn status are independent from each other.

As seen Table 4.15 and Table 4.16, Pearson Chi-square asymptotic significance value was found as 0,00 which is less than 0,05. Therefore, H0 is rejected which means that there is a relation between age and churn status. Results show that youth and older customers are not abandoning the company when it is compared to 27-40 age customer segment.

Table 4.15: Age-churn status test

| | | | churnstatus | | Total |
|---------|----------------------|----------------------|-------------|--------|--------|
| | | | 0 | 1 | |
| New Age | 27-40 | Count | 81 | 63 | 144 |
| | | Expected Count | 101,4 | 42,6 | 144,0 |
| | | % within New Age | 56,3% | 43,8% | 100,0% |
| | | % within churnstatus | 26,8% | 49,6% | 33,6% |
| | | % of Total | 18,9% | 14,7% | 33,6% |
| | 40+ | Count | 98 | 49 | 147 |
| | | Expected Count | 103,5 | 43,5 | 147,0 |
| | | % within New Age | 66,7% | 33,3% | 100,0% |
| | | % within churnstatus | 32,5% | 38,6% | 34,3% |
| | | % of Total | 22,8% | 11,4% | 34,3% |
| | 0-26 | Count | 123 | 15 | 138 |
| | | Expected Count | 97,1 | 40,9 | 138,0 |
| | | % within New Age | 89,1% | 10,9% | 100,0% |
| | | % within churnstatus | 40,7% | 11,8% | 32,2% |
| | | % of Total | 28,7% | 3,5% | 32,2% |
| Total | Count | 302 | 127 | 429 | |
| | Expected Count | 302,0 | 127,0 | 429,0 | |
| | % within New Age | 70,4% | 29,6% | 100,0% | |
| | % within churnstatus | 100,0% | 100,0% | 100,0% | |
| | % of Total | 70,4% | 29,6% | 100,0% | |

Table 4.16: Age-churn status Chi-square results

| Chi-Square Tests | | | Asymptotic Significance (2-sided) |
|-------------------------|---------------------|----|---|
| | Value | df | |
| Pearson Chi-Square | 38,050 ^a | 2 | ,000 |
| Likelihood Ratio | 41,820 | 2 | ,000 |
| N of Valid Cases | 429 | | |

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 40,85.

As seen Table 4.17 and Table 4.18, OOB minutes usage and churn status are dependent from each other.

Table 4.17: OOB usage-churn status test

New OOBUsagemin * churnstatus Crosstabulation

| | | new churnstatus | | Total | |
|------------------------------|----------|------------------------------|--------|--------|--------|
| | | 0 | 1 | | |
| New OOBUsagemin | 0,501-1 | Count | 61 | 41 | 102 |
| | | Expected Count | 71,8 | 30,2 | 102,0 |
| | | % within New OOBUsagemin | 59,8% | 40,2% | 100,0% |
| | | % within churnstatus | 20,2% | 32,3% | 23,8% |
| | | % of Total | 14,2% | 9,6% | 23,8% |
| | 0 | Count | 99 | 39 | 138 |
| | | Expected Count | 97,1 | 40,9 | 138,0 |
| | | % within New OOBUsagemin | 71,7% | 28,3% | 100,0% |
| | | % within churnstatus | 32,8% | 30,7% | 32,2% |
| | | % of Total | 23,1% | 9,1% | 32,2% |
| | 0,01-0,5 | Count | 142 | 47 | 189 |
| | | Expected Count | 133,0 | 56,0 | 189,0 |
| | | % within New OOBUsagemin2 | 75,1% | 24,9% | 100,0% |
| | | % within new churnstatus | 47,0% | 37,0% | 44,1% |
| | | % of Total | 33,1% | 11,0% | 44,1% |
| | Total | Count | 302 | 127 | 429 |
| Expected Count | | 302,0 | 127,0 | 429,0 | |
| % within New OOBUsagemin2 | | 70,4% | 29,6% | 100,0% | |
| % within new churnstatus | | 100,0% | 100,0% | 100,0% | |
| % of Total | | 70,4% | 29,6% | 100,0% | |

Table 4.18: OOB usage-churn status Chi-square results

| Chi-Square Tests | | | |
|-------------------------|--------------------|----|-----------------------------------|
| | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 7,645 ^a | 2 | ,022 |
| Likelihood Ratio | 7,412 | 2 | ,025 |
| N of Valid Cases | 429 | | |

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 30,20.

As seen in the tables below, Offnet usage rate and churn status are dependent from each other as a result of probability of Chi-Square test statistic which is less than %5. Test results indicate that customers who have high and medium amount of offnet usage rate willing to not abandon the company.

Table 4.19: Offnet usage rate-churn status test

| offnet Usage Rate * churnstatus Crosstabulation | | | | | |
|--|---------|--------------------------------|--------|--------|--------|
| | | new churnstatus | | | |
| | | | 0 | 1 | Total |
| New offnet Usage Rate | high | Count | 158 | 41 | 199 |
| | | Expected Count | 140,1 | 58,9 | 199,0 |
| | | % within New offnet Usage Rate | 79,4% | 20,6% | 100,0% |
| | | % within churnstatus | 52,3% | 32,3% | 46,4% |
| | | % of Total | 36,8% | 9,6% | 46,4% |
| | low | Count | 24 | 38 | 62 |
| | | Expected Count | 43,6 | 18,4 | 62,0 |
| | | % within New offnet Usage Rate | 38,7% | 61,3% | 100,0% |
| | | % within churnstatus | 7,9% | 29,9% | 14,5% |
| | | % of Total | 5,6% | 8,9% | 14,5% |
| | mediu m | Count | 120 | 48 | 168 |
| | | Expected Count | 118,3 | 49,7 | 168,0 |
| | | % within New offnet Usage Rate | 71,4% | 28,6% | 100,0% |
| | | % within churnstatus | 39,7% | 37,8% | 39,2% |
| | | % of Total | 28,0% | 11,2% | 39,2% |
| | Total | Count | 302 | 127 | 429 |
| Expected Count | | 302,0 | 127,0 | 429,0 | |
| % within New offnet Usage Rate | | 70,4% | 29,6% | 100,0% | |
| % within churnstatus | | 100,0% | 100,0% | 100,0% | |
| % of Total | | 70,4% | 29,6% | 100,0% | |

Table 4.20: Offnet usage rate-churn status Chi-square results

| Chi-Square Tests | | | |
|-------------------------|---------------------|----|---|
| | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 37,693 ^a | 2 | ,000 |
| Likelihood Ratio | 34,986 | 2 | ,000 |
| N of Valid Cases | 429 | | |

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 18,35.

As seen Table 4.21 and Table 4.22, Incoming/outgoing rate and churn status are independent from each other. Pearson Chi-Square asymptotic significance value which is found 0,00 is lower than 0,05. It means that H0 hypothesis is rejected and there is a relation between incoming/outgoing rate and churn status.

Table 4.21: Incoming/outgoing rate-churn status test

new Incomingoutgoing * churnstatus Crosstabulation

| | | new churnstatus | | Total |
|----------------------|-------------------------------|-----------------|--------|--------|
| | | 0 | 1 | |
| new Incomingoutgoing | Count | 20 | 5 | 25 |
| | Expected Count | 17,6 | 7,4 | 25,0 |
| | % within new Incomingoutgoing | 80,0% | 20,0% | 100,0% |
| | % within churnstatus | 6,6% | 3,9% | 5,8% |
| | % of Total | 4,7% | 1,2% | 5,8% |
| 0,501-1 | Count | 248 | 100 | 348 |
| | Expected Count | 245,0 | 103,0 | 348,0 |
| | % within new Incomingoutgoing | 71,3% | 28,7% | 100,0% |
| | % within churnstatus | 82,1% | 78,7% | 81,1% |
| | % of Total | 57,8% | 23,3% | 81,1% |
| 0 | Count | 0 | 7 | 7 |
| | Expected Count | 4,9 | 2,1 | 7,0 |
| | % within new Incomingoutgoing | 0,0% | 100,0% | 100,0% |
| | % within churnstatus | 0,0% | 5,5% | 1,6% |
| | % of Total | 0,0% | 1,6% | 1,6% |
| 0,01-0,5 | Count | 34 | 15 | 49 |
| | Expected Count | 34,5 | 14,5 | 49,0 |
| | % within new Incomingoutgoing | 69,4% | 30,6% | 100,0% |
| | % within churnstatus | 11,3% | 11,8% | 11,4% |
| | % of Total | 7,9% | 3,5% | 11,4% |
| Total | Count | 302 | 127 | 429 |
| | Expected Count | 302,0 | 127,0 | 429,0 |
| | % within new Incomingoutgoing | 70,4% | 29,6% | 100,0% |
| | % within churnstatus | 100,0% | 100,0% | 100,0% |
| | % of Total | 70,4% | 29,6% | 100,0% |

Table 4.22: Incoming/outgoing rate-churn status Chi-square results

Chi-Square Tests

| | Value | df | Asymptotic Significance (2-sided) |
|--------------------|---------------------|----|-----------------------------------|
| Pearson Chi-Square | 17,902 ^a | 3 | ,000 |
| Likelihood Ratio | 18,386 | 3 | ,000 |
| N of Valid Cases | 429 | | |

a. 2 cells (25,0%) have expected count less than 5.
The minimum expected count is 2,07.

As seen Table 4.23 and Table 4.24, Duration usage rate and churn status are independent from each other. Because, Probability of Pearson Chi-Square value is higher than %20. Duration usage rate does not change according to churn status.

Table 4.23: Duration usage rate-churn status test

new DurationUsageRate * churnstatus Crosstabulation

| | | new churnstatus | | Total |
|--------------------------|-----------------------------------|-----------------|--------|--------|
| | | 0 | 1 | |
| new DurationUsageRate | Count | 2 | 0 | 2 |
| | Expected Count | 1,4 | ,6 | 2,0 |
| | % within new DurationUsageRate | 100,0% | 0,0% | 100,0% |
| | % within churnstatus | 0,7% | 0,0% | 0,5% |
| | % of Total | 0,5% | 0,0% | 0,5% |
| 0,501-1 | Count | 150 | 65 | 215 |
| | Expected Count | 151,4 | 63,6 | 215,0 |
| | % within new DurationUsageRate | 69,8% | 30,2% | 100,0% |
| | % within churnstatus | 49,7% | 51,2% | 50,1% |
| | % of Total | 35,0% | 15,2% | 50,1% |
| 0 | Count | 8 | 3 | 11 |
| | Expected Count | 7,7 | 3,3 | 11,0 |
| | % within new DurationUsageRate | 72,7% | 27,3% | 100,0% |
| | % within churnstatus | 2,6% | 2,4% | 2,6% |
| | % of Total | 1,9% | 0,7% | 2,6% |
| 0,01-0,5 | Count | 142 | 59 | 201 |
| | Expected Count | 141,5 | 59,5 | 201,0 |
| | % within new DurationUsageRate | 70,6% | 29,4% | 100,0% |
| | % within churnstatus | 47,0% | 46,5% | 46,9% |
| | % of Total | 33,1% | 13,8% | 46,9% |
| Total | Count | 302 | 127 | 429 |
| | Expected Count | 302,0 | 127,0 | 429,0 |
| | % within new DurationUsageRate | 70,4% | 29,6% | 100,0% |
| | % within churnstatus | 100,0% | 100,0% | 100,0% |
| | % of Total | 70,4% | 29,6% | 100,0% |

Table 4.24: Duration usage rate-churn status Chi-square results

Chi-Square Tests

| | Value | df | Asymptotic Significance (2-sided) |
|--------------------|-------------------|----|-----------------------------------|
| Pearson Chi-Square | ,917 ^a | 3 | ,821 |
| Likelihood Ratio | 1,480 | 3 | ,687 |
| N of Valid Cases | 429 | | |

a. 3 cells (37,5%) have expected count less than 5. The minimum expected count is ,59.

Data usage rate and churn status are independent from each other and there is no relation between these variables as seen Table 4.25 and Table 4.26.

Table 4.25: Data usage rate-churn status test

new DataUsageRate * churnstatus Crosstabulation

| | | new churnstatus | | Total |
|-------------------|-----------------------------|-----------------|--------|--------|
| | | 0 | 1 | |
| new DataUsageRate | Count | 1 | 0 | 1 |
| | Expected Count | ,7 | ,3 | 1,0 |
| | % within new DataUsageRate2 | 100,0% | 0,0% | 100,0% |
| | % within churnstatus | 0,3% | 0,0% | 0,2% |
| | % of Total | 0,2% | 0,0% | 0,2% |
| | | | | |
| 0,501-1 | Count | 194 | 69 | 263 |
| | Expected Count | 185,1 | 77,9 | 263,0 |
| | % within new DataUsageRate | 73,8% | 26,2% | 100,0% |
| | % within churnstatus | 64,2% | 54,3% | 61,3% |
| | % of Total | 45,2% | 16,1% | 61,3% |
| | | | | |
| 0 | Count | 42 | 27 | 69 |
| | Expected Count | 48,6 | 20,4 | 69,0 |
| | % within new DataUsageRate | 60,9% | 39,1% | 100,0% |
| | % within churnstatus | 13,9% | 21,3% | 16,1% |
| | % of Total | 9,8% | 6,3% | 16,1% |
| | | | | |
| 0,01-0,5 | Count | 65 | 31 | 96 |
| | Expected Count | 67,6 | 28,4 | 96,0 |
| | % within new DataUsageRate | 67,7% | 32,3% | 100,0% |
| | % within churnstatus | 21,5% | 24,4% | 22,4% |
| | % of Total | 15,2% | 7,2% | 22,4% |
| | | | | |
| Total | Count | 302 | 127 | 429 |
| | Expected Count | 302,0 | 127,0 | 429,0 |
| | % within new DataUsageRate | 70,4% | 29,6% | 100,0% |
| | % within churnstatus | 100,0% | 100,0% | 100,0% |
| | % of Total | 70,4% | 29,6% | 100,0% |
| | | | | |

Table 4.26: Data usage rate-churn status Chi-square results

| Chi-Square Tests | | | |
|-------------------------|--------------------|----|---|
| | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 5,190 ^a | 3 | ,158 |
| Likelihood Ratio | 5,347 | 3 | ,148 |
| N of Valid Cases | 429 | | |

a. 2 cells (25,0%) have expected count less than 5. The minimum expected count is ,30.

SMS usage rate and churn status do not have a relation as seen Table 4.27 and Table 4.28. Probability of Pearson chi-square value which is 0,298 is higher than 0,05. SMS usage rate does not change according to churn status.

Table 4.27: SMS usage rate-churn status test

New SMS Usage Rate * churn status Crosstabulation

| | | churn status | | Total | |
|-----------------------|--------------------------------|--------------------------------|---------|--------|--------|
| | | nonchurner | churner | | |
| New SMS Usage Rate | Count | 0 | 1 | 1 | |
| | % within New SMS Usage Rate | 0,0% | 100,0% | 100,0% | |
| | % within churn status | 0,0% | 0,8% | 0,2% | |
| | % of Total | 0,0% | 0,2% | 0,2% | |
| | high | Count | 43 | 17 | 60 |
| | | % within New SMS Usage Rate | 71,7% | 28,3% | 100,0% |
| | | % within churn status | 14,2% | 13,4% | 14,0% |
| | | % of Total | 10,0% | 4,0% | 14,0% |
| | low | Count | 259 | 109 | 368 |
| | | % within New SMS Usage Rate | 70,4% | 29,6% | 100,0% |
| | | % within churn status | 85,8% | 85,8% | 85,8% |
| | | % of Total | 60,4% | 25,4% | 85,8% |
| Total | Count | 302 | 127 | 429 | |
| | % within New SMS Usage Rate | 70,4% | 29,6% | 100,0% | |
| | % within churn status | 100,0% | 100,0% | 100,0% | |
| | % of Total | 70,4% | 29,6% | 100,0% | |

Table 4.28: SMS usage rate-churn status Chi-square results

| Chi-Square Tests | | | |
|---------------------|--------------------|----|-----------------------------------|
| | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 2,424 ^a | 2 | ,298 |
| Likelihood Ratio | 2,481 | 2 | ,289 |
| McNemar-Bowker Test | . | . | . ^b |
| N of Valid Cases | 429 | | |

a. 2 cells (33,3%) have expected count less than 5. The minimum expected count is ,30.

b. Computed only for a P x P table, where P must be greater than 1.

As seen Table 4.29 and Table 4.31 below, Pearson Chi-Square values of migration and tenure which are 0,00 and 0,003 respectively are lower than 0,05. Therefore, these variables have a relation between churn status separately. Results in Table 4.29 shows that loyal customers who are 25+ months tenure willing to stay in the existing carrier' customer base. In addition, Table 4.31 indicate that if customers have low or medium rate of migration, they are more loyal to existing carrier.

Table 4.29: Tenure -churn status test

New Tenure *churnstatus Crosstabulation

| | | new churnstatus | | Total | |
|------------|----------------------|----------------------|--------|--------|--------|
| | | 0 | 1 | | |
| New Tenure | 25+ | Count | 157 | 25 | 182 |
| | | Expected Count | 128,1 | 53,9 | 182,0 |
| | | % within New Tenure | 86,3% | 13,7% | 100,0% |
| | | % within churnstatus | 52,0% | 19,7% | 42,4% |
| | | % of Total | 36,6% | 5,8% | 42,4% |
| | 0-12 | Count | 52 | 68 | 120 |
| | | Expected Count | 84,5 | 35,5 | 120,0 |
| | | % within New Tenure | 43,3% | 56,7% | 100,0% |
| | | % within churnstatus | 17,2% | 53,5% | 28,0% |
| | | % of Total | 12,1% | 15,9% | 28,0% |
| | 13-25 | Count | 93 | 34 | 127 |
| | | Expected Count | 89,4 | 37,6 | 127,0 |
| | | % within New Tenure | 73,2% | 26,8% | 100,0% |
| | | % within churnstatus | 30,8% | 26,8% | 29,6% |
| | | % of Total | 21,7% | 7,9% | 29,6% |
| Total | Count | 302 | 127 | 429 | |
| | Expected Count | 302,0 | 127,0 | 429,0 | |
| | % within New Tenure | 70,4% | 29,6% | 100,0% | |
| | % within churnstatus | 100,0% | 100,0% | 100,0% | |
| | % of Total | 70,4% | 29,6% | 100,0% | |

Table 4.30: Tenure-churn status Chi-square results

| Chi-Square Tests | | | |
|-------------------------|---------------------|----|-----------------------------------|
| | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 64,650 ^a | 2 | ,000 |
| Likelihood Ratio | 63,772 | 2 | ,000 |
| N of Valid Cases | 429 | | |

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 35,52.

Table 4.31: Migration -churn status test

New Migration * churn status Crosstabulation

| | | churn status | | Total | |
|---------------|------------------------|------------------------|---------|--------|--------|
| | | nonchurner | churner | | |
| New Migration | high | Count | 8 | 2 | 10 |
| | | % within New Migration | 80,0% | 20,0% | 100,0% |
| | | % within churn status | 2,6% | 1,6% | 2,3% |
| | | % of Total | 1,9% | 0,5% | 2,3% |
| | Low | Count | 110 | 69 | 179 |
| | | % within New Migration | 61,5% | 38,5% | 100,0% |
| | | % within churn status | 36,4% | 54,3% | 41,7% |
| | | % of Total | 25,6% | 16,1% | 41,7% |
| | medium | Count | 184 | 56 | 240 |
| | | % within New Migration | 76,7% | 23,3% | 100,0% |
| | | % within churn status | 60,9% | 44,1% | 55,9% |
| | | % of Total | 42,9% | 13,1% | 55,9% |
| Total | Count | 302 | 127 | 429 | |
| | % within New Migration | 70,4% | 29,6% | 100,0% | |
| | % within churn status | 100,0% | 100,0% | 100,0% | |
| | % of Total | 70,4% | 29,6% | 100,0% | |

Table 4.32: Migration-churn status Chi-square results

Chi-Square Tests

| | Value | df | Asymptotic Significance (2-sided) |
|---------------------|---------------------|----|-----------------------------------|
| Pearson Chi-Square | 11,841 ^a | 2 | ,003 |
| Likelihood Ratio | 11,758 | 2 | ,003 |
| McNemar-Bowker Test | . | . | . ^b |
| N of Valid Cases | 429 | | |

a. 1 cells (16,7%) have expected count less than 5. The minimum expected count is 2,96.

b. Computed only for a P x P table, where P must be greater than 1.

Variables named OOB Data Usage Rate, OOB Usage Rate, Port-in, Port-out and # of Top-up (annual) were analyzed using Mc-Nemar test in SPSS program. Churner and non-churner dataset are separately analyzed unlike Chi-Square test.

As seen Table 4.33 and Table 4.34, there is a relation between OOB Data Usage Rate and OOB SMS Usage Rate using non- churner dataset. The results are indicated that if the rate of OOB Data Usage is low, OOB SMS usage rate is also low.

Table 4.33: OOB Data usage - OOB SMS usage test using non-churner dataset

New OOBDData * New OOBSMSUsage Crosstabulation

| | | | New OOBSMSUsage | | Total |
|--------------|-------|--------------------------|-----------------|--------|--------|
| | | | low | high | |
| New OOBDData | low | Count | 191 | 38 | 229 |
| | | % within New OOBDData | 83,4% | 16,6% | 100,0% |
| | | % within New OOBSMSUsage | 73,7% | 88,4% | 75,8% |
| | | % of Total | 63,2% | 12,6% | 75,8% |
| New OOBDData | high | Count | 68 | 5 | 73 |
| | | % within New OOBDData | 93,2% | 6,8% | 100,0% |
| | | % within New OOBSMSUsage | 26,3% | 11,6% | 24,2% |
| | | % of Total | 22,5% | 1,7% | 24,2% |
| Total | Total | Count | 259 | 43 | 302 |
| | | % within New OOBDData | 85,8% | 14,2% | 100,0% |
| | | % within New OOBSMSUsage | 100,0% | 100,0% | 100,0% |
| | | % of Total | 85,8% | 14,2% | 100,0% |

Table 4.34: OOB Data usage rate- OOB SMS usage rate results using non-churner dataset

| Chi-Square Tests | | | | | |
|------------------------------------|--------------------|----|-----------------------|----------------------|----------------------|
| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | 4,304 ^a | 1 | ,038 | | |
| Continuity Correction ^b | 3,543 | 1 | ,060 | | |
| Likelihood Ratio | 4,918 | 1 | ,027 | | |
| Fisher's Exact Test | | | | ,053 | ,025 |
| Linear-by-Linear Association | 4,290 | 1 | ,038 | | |
| McNemar Test | | | | ,005 ^c | |
| N of Valid Cases | 302 | | | | |

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 10,39.

b. Computed only for a 2x2 table

c. Binomial distribution used.

As seen Table 4.35 and Table 4.36, variables named port-in and port-out have a relation using non-churner dataset. The test result shows that when port-in and port-out are both “0”, customers are willing to stay in the company customer base.

Table 4.35: Port-in- port-out test using non-churner dataset

| New Portin * New port out Crosstabulation | | | | | |
|---|----------|-----------------------|---------------|----------|--------|
| | | | New4 port out | | Total |
| | | | negative | positive | |
| New Portin | negative | Count | 247 | 6 | 253 |
| | | % within New Portin | 97,6% | 2,4% | 100,0% |
| | | % within New port out | 84,0% | 75,0% | 83,8% |
| | | % of Total | 81,8% | 2,0% | 83,8% |
| positive | Count | Count | 47 | 2 | 49 |
| | | % within New Portin | 95,9% | 4,1% | 100,0% |
| | | % within New port out | 16,0% | 25,0% | 16,2% |
| | | % of Total | 15,6% | ,7% | 16,2% |
| Total | Count | Count | 294 | 8 | 302 |
| | | % within New Portin | 97,4% | 2,6% | 100,0% |
| | | % within New port out | 100,0% | 100,0% | 100,0% |
| | | % of Total | 97,4% | 2,6% | 100,0% |

Table 4.36: Port-in- port-out results using non-churner dataset

| Chi-Square Tests | | | | | |
|------------------------------------|-------------------|----|-----------------------|----------------------|----------------------|
| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | ,466 ^a | 1 | ,495 | | |
| Continuity Correction ^b | ,039 | 1 | ,844 | | |
| Likelihood Ratio | ,414 | 1 | ,520 | | |
| Fisher's Exact Test | | | | ,620 | ,382 |
| Linear-by-Linear Association | ,464 | 1 | ,496 | | |
| McNemar Test | | | | ,000 ^c | |
| N of Valid Cases | 302 | | | | |

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 1,30.

b. Computed only for a 2x2 table

c. Binomial distribution used.

As seen Table 4.37 and Table 4.38, OOB Data Usage Rate and # of Top-up (annual) are dependent from each other using non-churner dataset. The results are indicated that customers who do not top-up are not exceed their data tariffs.

Table 4.37: OOB data usage- top-up test using non-churner dataset

| New OOBData * New topup Crosstabulation | | | | | |
|---|------|----------------------|-----------|--------|--------|
| | | | New topup | | |
| | | | ,00 | 1,00 | Total |
| New OOBData | low | Count | 138 | 91 | 229 |
| | | % within New OOBData | 60,3% | 39,7% | 100,0% |
| | | % within New topup | 74,6% | 77,8% | 75,8% |
| | | % of Total | 45,7% | 30,1% | 75,8% |
| | high | Count | 47 | 26 | 73 |
| | | % within New OOBData | 64,4% | 35,6% | 100,0% |
| | | % within New topup | 25,4% | 22,2% | 24,2% |
| | | % of Total | 15,6% | 8,6% | 24,2% |
| Total | | Count | 185 | 117 | 302 |
| | | % within New OOBData | 61,3% | 38,7% | 100,0% |
| | | % within New topup | 100,0% | 100,0% | 100,0% |
| | | % of Total | 61,3% | 38,7% | 100,0% |

Table 4.38: OOB data usage rate- top-up results using non-churner dataset

| Chi-Square Tests | | | | | |
|------------------------------------|-------------------|----|-----------------------|----------------------|----------------------|
| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | ,396 ^a | 1 | ,529 | | |
| Continuity Correction ^b | ,242 | 1 | ,623 | | |
| Likelihood Ratio | ,399 | 1 | ,528 | | |
| Fisher's Exact Test | | | | ,582 | ,313 |
| Linear-by-Linear Association | ,395 | 1 | ,530 | | |
| McNemar Test | | | | ,000 ^c | |
| N of Valid Cases | 302 | | | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 28,28.

b. Computed only for a 2x2 table

c. Binomial distribution used.

As seen Table 4.39 and 4.40, OOB Data Usage Rate and and OOB SMS Usage rate are independent from each other using churner data. There is no relation between two variables.

Table 4.39: OOB Data usage- OOB SMS usage test using churner dataset

| New OOB SMS Usage * New OOB Data Crosstabulation | | | | | |
|--|-------|----------------------------|--------------|--------|--------|
| | | | New OOB Data | | |
| | | | low | high | Total |
| New OOB SMS Usage | low | Count | 91 | 18 | 109 |
| | | % within New OOB SMS Usage | 83,5% | 16,5% | 100,0% |
| | | % within New OOB Data | 88,3% | 78,3% | 86,5% |
| | | % of Total | 72,2% | 14,3% | 86,5% |
| | high | Count | 12 | 5 | 17 |
| | | % within New OOB SMS Usage | 70,6% | 29,4% | 100,0% |
| | | % within New OOB Data | 11,7% | 21,7% | 13,5% |
| | | % of Total | 9,5% | 4,0% | 13,5% |
| | Total | Count | 103 | 23 | 126 |
| | | % within New OOB SMS Usage | 81,7% | 18,3% | 100,0% |
| | | % within New OOB Data | 100,0% | 100,0% | 100,0% |
| | | % of Total | 81,7% | 18,3% | 100,0% |

Table 4.40: OOB Data usage- OOB SMS usage results using churner dataset

Chi-Square Tests

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|------------------------------------|--------------------|----|--------------------------|-------------------------|-------------------------|
| Pearson Chi-Square | 1,640 ^a | 1 | ,200 | | |
| Continuity Correction ^b | ,889 | 1 | ,346 | | |
| Likelihood Ratio | 1,475 | 1 | ,225 | | |
| Fisher's Exact Test | | | | ,196 | ,171 |
| Linear-by-Linear Association | 1,627 | 1 | ,202 | | |
| McNemar Test | | | | ,362 ^c | |
| N of Valid Cases | 126 | | | | |

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 3,10.

b. Computed only for a 2x2 table

c. Binomial distribution used.

As seen Table 4.41 and Table 4.42 below, port-in and port-out are dependent from each other using churner data. There is a relation between two variables. The test result indicates that when port-in and port-out are both “0”, customers are willing to stay in the company customer base like on non-churner data as seen Table 4.35.

Table 4.41: Port-in- port-out test using churner dataset

New Portin * New port out Crosstabulation

| | | | New port out | | |
|------------|----------|-----------------------|--------------|----------|--------|
| | | | negative | positive | Total |
| New Portin | negative | Count | 91 | 0 | 91 |
| | | % within New Portin | 100,0% | ,0% | 100,0% |
| | | % within New port out | 74,6% | ,0% | 71,7% |
| | | % of Total | 71,7% | ,0% | 71,7% |
| positive | Count | Count | 31 | 5 | 36 |
| | | % within New Portin | 86,1% | 13,9% | 100,0% |
| | | % within New port out | 25,4% | 100,0% | 28,3% |
| | | % of Total | 24,4% | 3,9% | 28,3% |
| Total | Count | Count | 122 | 5 | 127 |
| | | % within New Portin | 96,1% | 3,9% | 100,0% |
| | | % within New port out | 100,0% | 100,0% | 100,0% |
| | | % of Total | 96,1% | 3,9% | 100,0% |

Table 4.42: Port-in- port-out results using churner dataset

| Chi-Square Tests | | | | | |
|------------------------------------|---------------------|----|--------------------------|-------------------------|-------------------------|
| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | 13,157 ^a | 1 | ,000 | | |
| Continuity Correction ^b | 9,741 | 1 | ,002 | | |
| Likelihood Ratio | 13,136 | 1 | ,000 | | |
| Fisher's Exact Test | | | | ,001 | ,001 |
| Linear-by-Linear Association | 13,053 | 1 | ,000 | | |
| McNemar Test | | | | ,000 ^c | |
| N of Valid Cases | 127 | | | | |

a. 2 cells (50,0%) have expected count less than 5. The minimum expected count is 1,42.

b. Computed only for a 2x2 table

c. Binomial distribution used.

As seen Table 4.43 and Table 4.44, OOB Data Usage Rate and Top-up (annual) are independent from each other using churner data. There is no relation between two variables using Pearson Chi –Square test.

Table 4.43: OOB data usage rate- top-up test using churner dataset

| New OOBData * New Topup Crosstabulation | | | | | |
|---|----------------------|----------------------|-----------|--------|--------|
| | | | New Topup | | |
| | | | High | low | Total |
| New OOBData | low | Count | 26 | 78 | 104 |
| | | % within New OOBData | 25,0% | 75,0% | 100,0% |
| | | % within New Topup | 76,5% | 83,9% | 81,9% |
| | | % of Total | 20,5% | 61,4% | 81,9% |
| high | Count | 8 | 15 | 23 | |
| | % within New OOBData | 34,8% | 65,2% | 100,0% | |
| | % within New Topup | 23,5% | 16,1% | 18,1% | |
| | % of Total | 6,3% | 11,8% | 18,1% | |
| Total | Count | 34 | 93 | 127 | |
| | % within New OOBData | 26,8% | 73,2% | 100,0% | |
| | % within New Topup | 100,0% | 100,0% | 100,0% | |
| | % of Total | 26,8% | 73,2% | 100,0% | |

Table 4.44: OOB data usage rate- top-up results using churner dataset

| Chi-Square Tests | | | | | |
|------------------------------------|-------------------|----|--------------------------|-------------------------|-------------------------|
| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | ,919 ^a | 1 | ,338 | | |
| Continuity Correction ^b | ,488 | 1 | ,485 | | |
| Likelihood Ratio | ,882 | 1 | ,348 | | |
| Fisher's Exact Test | | | | ,435 | ,238 |
| McNemar Test | | | | . ^c | |
| N of Valid Cases | 127 | | | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6,16.

b. Computed only for a 2x2 table

c. Both variables must have identical values of categories.

4.3 Churn Prediction Methods

Previous studies indicated that churn prediction in telecommunication sector primarily is calculated by using Artificial Neural Network, Decision Trees, Logistic Regression, Support Vector Machine, Random Forest, Naïve Bayes, Gradient Boosting and etc.

Altas D. and Gülpinar G. (2013) identify the factors of the churn prediction, studying on Artificial Neural Network and find patterns of churn in Turkish telecommunication market. Also, Nie et al. (2011) study compare that credit card churn prediction by logistic regression and decision tree algorithms using a data set collected from Chinese Bank. On the contrary, Lemmens and Croux (2006) study explore that bagging and boosting classification techniques for implementing to telecommunication industry churn prediction.

In the following section, commonly used methods explained briefly.

4.3.1 Artificial neural network

Artificial Neural Network is a common method is being used not only for prediction of customer churn but also in complex problem solving. Neural Networks can be based on hardware which represent the physical components or software also can be used several learning algorithms. In some researches, neural networks have better performance than the prediction of churn compare with Decision Tree (Au et al.,)

4.3.2 Support vector machine

Support Vector Machine Algorithm is generally used for analyzing data and patterns of the customers. This method is known as functional risk hedger. In some cases, Support Vector Machine method provides more accurate results depending on data variations compared to Decision Tree and ANN.

4.3.3 Naïve bayes

Naïve Bayes is called as a simple probabilistic classifier using Bayes theorem. This classifier helps to specific characteristic of a class (e.g churn) relation between other features in terms of presence or absence. This method is widely used in telecommunication industry for churn prediction.

4.3.4 Decision tree algorithm

In this paper, customer churn is shaped by Decision Tree algorithm, which is also used in literature. Nie et al. (2011) explained that decision tree is a learning technique that organizes variables from a sample dataset in a functional model consist of nodes and ramifications. It is easy to understand that the form of output of the decision tree thanks to its hierarchical structure. According to Lorena & Carvalho (2007), Decision Tree algorithm can be used numerical and categorical data for building models. Target variable values are consist of nodes and its aggregated values. Each node has a probability for each class in a decision tree. Also, this method is commonly used in specific functions like supplier choosing (Wu, 2009) or churn of internet company users (Nie, Zhang, Li, & Shi, 2006).

In the light of these researches, Decision Tree method is not widely used in telecommunication industry. Nevertheless, this method has meaningful structure as interpreting and implementing the results rapidly. For this reason, in this thesis Decision Tree algorithm is selected for data analysis.



5. ANALYSIS AND FINDINGS

5.1 Implementation

Firstly, training dataset is run in SPSS V22.0 using Decision tree model Chaid algorithm effectively. In this algorithm, as seen below Table 5.1, churn rate probability of prepaid customers using non-churner data is found %95,7, and possibility of being churn of the customer using churner data is found %70,9 which is acceptable. This result means that if the model labelled the customer as non-churner, the customer will not abandon the company with a probability of %95,7. Also general probability of churn rate is predicted as %88,3 by the model for this sampling.

The detail results are displayed in Figure 5.1 for this algorithm. Algorithm is selected most meaningful and accurate variable in first ramification, which is called offnet usage rate. Offnet usage rate is basically defined as making other operators' customers call. In Node 1 which is offnet usage rate lower than 0,0129, churn probability of customers is %65,9. That means, if customers make few call to other operators, they willing to abandon the existing carrier. It conflicts with literature, because the training dataset is selected 3 months in advance. It is shows that, it would be better to be selected earlier for accuracy of behavioral data. The results can be different if the dataset choose 6 or 9 months before the churn flag. In addition, prepaid customers does not have a commitment compare to postpaid customers. As seen Node 2 and Node 3 that if offnet usage rate between 0,0129 and 0,977, customer churn rate %44,2 and %7 respectively. Node 4 shows that making calls to other operators in %97 rate, customer churn probability increases to %79,1 which is in line with previous researches.

Node 1 samples are ramified to Node 5 and Node 6 using OOB (Out-of-Bundle) minutes usage defined that exceeded tariff bundle minutes. If OOB minutes usage lower than 0,0158, churner customer probability reaches to %93,3. This node can be eliminated. Because, these group of customers cannot have a tariff and all usages could be labeled as OOB by the system. Also, in Node 9, churn probability increased

compare to Node 7. It shows that, prepaid customers cannot manage their minutes usage and they preferred to abandon the company. In Node 9, churn probability decreased again. Because, prepaid customers can exceed their tariff until the end of their credit.

Table 5.1: First churn probability chart

| Classification | | | |
|--------------------|-------------|---------|-----------------|
| Observed | Predicted | | |
| | 0 NON CHURN | 1 CHURN | Percent Correct |
| 0 NON CHURN | 289 | 13 | 95,7% |
| 1 CHURN | 37 | 90 | 70,9% |
| Overall Percentage | 76,0% | 24,0% | 88,3% |

Growing Method: CHAID
 Dependent Variable: CHURN

Secondly, training dataset was run again in SPSS V22.0 using decision tree model CRT algorithm for comparing the two results. In this algorithm, as seen below Table 5.2, churn rate probability of prepaid customers using non-churner data is found to be %95,7, and possibility of being churn of the customer using churner data is found to be %23,6 which is not acceptable. This result confirms the hypothesis that churner data is not accurate and it should be selected in historical data from 6 or 9 months before.

Table 5.2: Second churn probability chart

| Classification | | | |
|--------------------|-------------|---------|-----------------|
| Observed | Predicted | | |
| | 0 NON CHURN | 1 CHURN | Percent Correct |
| 0 NON CHURN | 289 | 13 | 95,7% |
| 1 CHURN | 97 | 30 | 23,6% |
| Overall Percentage | 90,0% | 10,0% | 74,4% |

Growing Method: CRT
 Dependent Variable: CHURN

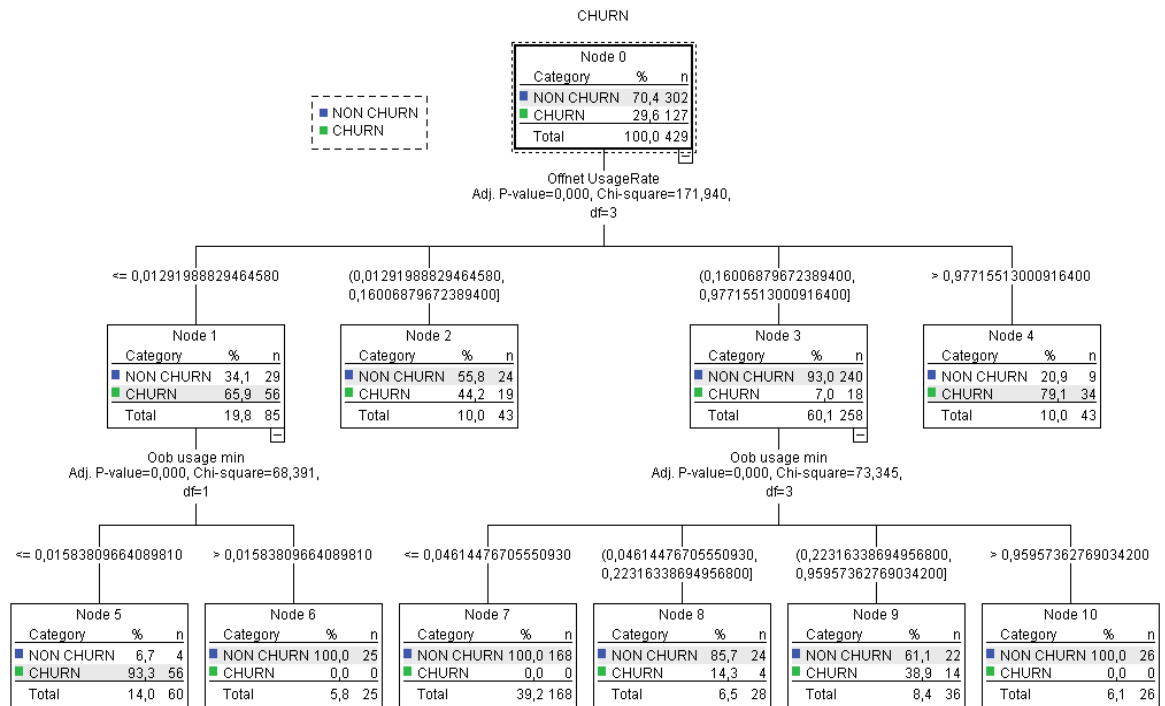


Figure 5.1: First result of churn probability.

As seen in Figure 5.2, customer tenure and age are another important variables in addition to offnet usage and OOB usage minutes. Customer tenure means that the spent months in the carrier base as explained Table 4.2.

First ramification shows that customer tenure and churn probability are inversely proportional. More than 27 months tenured customers are more loyal to carrier with % 85,8 probability which is an expected result.

Second ramification explains that youth and older customers are more loyal than the others. Group of 25-34 and 34-42 ages act similar in terms of churn movement. These age groups need different retention activity for staying in the carriers' customer base.

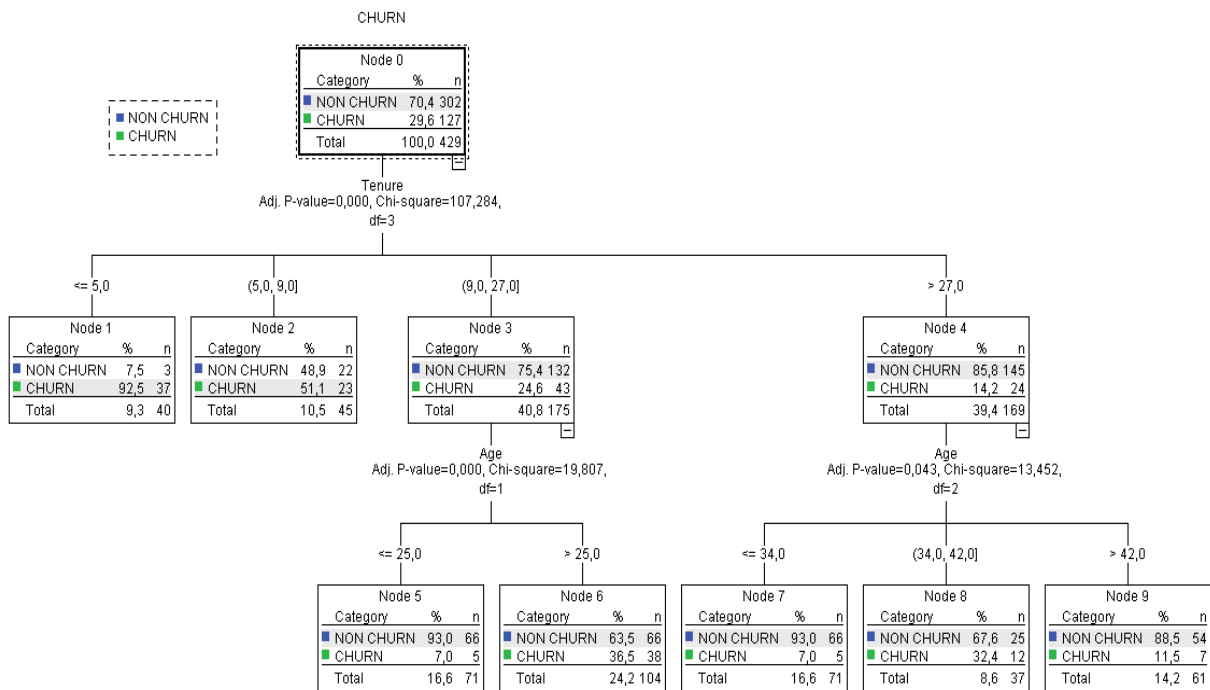


Figure 5.2: Second result of churn probability.

5.2 Interpretation of the Findings

Churn flagging time in the systems is an important finding for this thesis for many reasons:

- Customers who flagged as churner in 9 months after last top-up get damaged the non-churner dataset accuracy and probability of churn is decreasing.
- Churner customer probability are affected on other customers who actually gone from carrier before taking churn flag.

Retention activities are planned by customer relationship management depending on churn probability results. Misidentified customers as non-churner or churner have an extra retention cost to the companies. For managing this cost, dataset should be taken from the system in different periods (e.g. 3,6,9 months after last top-up) for comparing the movements.

Out of bundle minutes usage is another important parameter for churn probability. Although prepaid customers do not face with bill shock compare with postpaid customers, they can be feel “pissed off” the company about decreasing their credits because of exceeding their tariff. Because of that, they can abandon the company as seen in test results above. Under this circumstance, reacquisition rates might decrease. Some studies show (Kumar et al., 2015) that positive first-lifetime experience directly affects the acceptance of win-back offers.

Offnet usage rate is another important variable for prepaid churn prediction. As mentioned in the first sections, social network of the customers give us a hint about customer movement to other carriers. According to Oliver (1999), analyses show that the customer satisfaction is a key factor for creating loyalty, but loyalty is less important compared to other factors which are personal determinism and social network. Also, Dasgupta et al. (2008) found important results using Social Network Analysis techniques called Call Data Records which also can be named social ties in the literature, belong to customers effected to churn significantly in telecom industry. As managerial implications, every way use content in terms of tariff minutes can decrease churn rate from the company. Also, most churned months can be detected with further analysis. In these months, tariff usages can be investigated in depth. In the light of in-depth analysis, different proposals including minutes or data can be gifted to the customer for win them back.

Another significant parameter, customer tenure affects the customer churn probability inversely proportional. In line with this finding, Kumar et al. (2015) indicated that customer tenure has positive effect on churn rate calculation and profitability in industries.

5.3 Limitation of the Research

It should be noted that this study was run on a specific dataset and results could change depending on varying datasets and with different variables.

For future research; using historical data older than 3 months period will be more accurate for predicting prepaid customer churn. Also, as a diagnosis for defining prepaid churners in the company system, top-up period correlation is an important factor for behavioral data accuracy.



6. CONCLUSION

Telecommunication sector continuously enlarged their businesses in years. Therefore, competition between companies is increased and they tried to find a way to expand their market share. Churn prediction methods are started to be used frequently for keeping existing customers in company base.

This study represents a research on churn prediction methods and interpretation of variables in churn calculation. In this study, decision tree model using SPSS V22.0 program has been implemented to a sample dataset gathered from a telecommunication company in Turkey for forecasting customer churn behavior. Variable correlation which are defined in Section 4, and relation among them were analyzed deeply using Pearson Chi-Square and Mc Nemar test. It has been found that Duration usage rate, incoming/outgoing usage rate, SMS usage rate and data usage rate are not dependent from churn status respectively. In addition, OOB data usage rate do not have a relation between and OOB SMS usage rate or # of top-up respectively when analyzed using Mc Nemar analysis.

Substantial insights regarding reasons of prepaid customer churn are found so that retention activities can planned depending on these findings as suggested in Section 5.2. According to the results, the variables; offnet usage minutes, out of bundle usage, customer tenure and age are directly affecting the customer churn compared to other variables that are data/SMS usage, out of bundle data/SMS usage, incoming/outgoing minutes, number of migration, top-up, port-in and port-out.

It has been noticed in the findings that abandonment date is not known exactly because of the nature of the prepaid customers. Churner samples are not examined most correctly. It is selected 3 months before churn flag in this thesis. But it is proven that selection of 3 months before is not enough for analyzing behavior of the customer. Most of the sample data is eliminated as explained previous section because of no movement in selected month.



REFERENCES

- Ahn, H.J., Pil Han, S. and Seop Lee, Y.** (2006). Customer churn analysis: churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry”, *Telecommunications Policy*, Vol. 30, pp. 552-68.
- Allenby, G., Leone, R. and Jen, L.** (1999). A Dynamic Model of Purchase Timing with Application to Direct Marketing. *Journal of the American Statistical Association*, 94(446), p.365.
- Altas, D., Gülpinar, V.,** (2013). Customer Churn Analysis through Artificial Neural Networks in Turkish Telecommunications Market, *International Journal of Economic Perspectives*; Famagusta, Vol. 7, Iss. 4, (2013): 63-80.
- Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A. and Huang, K.** (2017). Customer churn prediction in the telecommunication sector using a rough set approach, *Neurocomputing*, 237, pp.242-254.
- Aydin K.** (2004). Üniversite öğrencilerinin cep telefonu kullanımı ve GSM operatörü tercihleri üzerine bir çalışma, *Suleyman Demirel Üniversitesi IIBF Dergisi*, Vol. 9 (1), pp. 149- 164.
- Bolton, R.** (1998). A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction, *Marketing Science*, 17(1), pp.45-65.
- Bolton, R.N., Lemon, K.N. and Verhoeff, P.C.** (2004). The theoretical underpinnings of customer asset management: a framework and propositions for future research, *Journal of the Academy of Marketing Science*, Vol. 32, pp. 271-92.
- Bonfrer, A., Knox, G., Eliashberg, J. and Chiang, J.** (2010). A First-Passage Time Model for Predicting Inactivity in a Contractual Setting, *SSRN Electronic Journal*.

- Buckinx, W., & Van den Poel, D.** (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting, *European Journal of Operational Research*, Vol. 164(1), pp. 252–268.
- Chandrashek, M., Rotte, K., Tax, S.S. and Grewal, R.** (2007). Satisfaction strength and customer loyalty”, *Journal of Marketing Research*, Vol. 44, pp. 153-63.
- Coussement K., Lessmann S., Verstraeten G.** (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry, *Decision Support Systems*, Vol. 95, pp. 27-36.
- Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanava, A. and Joshi, A.** (2008). Social ties and their relevance to churn in mobile telecom networks”, *EDBT’08: Proceedings of the 11th International Conference on Extending Database Technology*, New York, NY, USA, ACM, pp. 668-77.
- Dover, H.F. and Merthi, B.P.S.** (2006). Asymmetric effects of dynamic usage behavior on duration in subscription-based online service, *Journal of Interactive Marketing*, Vol. 20, pp. 5-15.
- Eui-Bang Lee, Jinwha Kim, Sang-Gun Lee** (2017). Predicting customer churn in mobile industry using data mining technology, *Industrial Management & Data Systems*, Vol. 117 Issue: 1, pp.90-109.
- Gale B.** (1992). Monitoring customer satisfaction and market-perceived quality. *Worth repeating series (vol. 922CSO1)*”, Chicago: American Marketing Association.
- Ganesh, J., Arnold, M.J. and Reinolds, K.E.** (2000). Understanding the customer base of service providers: an examination of the difference between switchers and stayers, *Journal of Marketing*, Vol. 64, pp. 65-87.
- Geetha, M. and Abitha Kumari, J.** (2012). Analysis of churn behavior of consumers in Indian telecom sector. *Journal of Indian Business Research*, 4(1), pp.24-35.
- Gerpott T., Rams W., & Schindler A.** (2001). Customer retention, loyalty and satisfaction in the German mobile cellular telecommunications market, *Telecommunications Policy*, Vol. 25(4), pp. 249–269.

- Giovanis, A. and Athanasopoulou, P.** (2018). Consumer-brand relationships and brand loyalty in technology-mediated services. *Journal of Retailing and Consumer Services*, 40, pp.287-294.
- Hughes, A.M.** (2005). *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable Customer-based Marketing Program*, McGraw-Hill, New York, NY.
- Inoue, Y., Funk, D. and McDonald, H.** (2017). Predicting behavioral loyalty through corporate social responsibility: The mediating role of involvement and commitment”, *Journal of Business Research*, 75, pp.46-56.
- Kamakura, W.A., Wedel, M., Rosa, F.D. and Mazzon, J.A.** (2003). Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction, *International Journal of Research in Marketing*, Vol. 20, pp. 45-65.
- Karacuka, M., Haucap, J. and Heimeshoff, U.** (2011). Competition in Turkish mobile telecommunications markets: Price elasticities and network substitution. *Telecommunications Policy*, 35(2), pp.202-210.
- Khodabandehlou, S. and Zivari Rahman, M.** (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*, 19(1/2), pp.65-93.
- Kim, H. and Yoon, C.** (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market, *Telecommunications Policy*, 28(9-10), pp.751-765.
- Kim M., Park M., & Jeong D.** (2004). The effects of customer satisfaction and switching barriers on customer loyalty in Korean mobile telecommunication services, *Telecommunications Policy*, Vol. 28(2), pp. 145–159.
- Kim Y. S., Moon S.** (2012). Measuring the success of retention management models built on churn probability, retention probability, and expected yearly revenues, *Expert Systems and Applications*, Vol. 39 Issue 14, pp. 11718-11727
- Kisioglu, P. and Topcu, Y.** (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), pp.7151-7157.

- Kumar, V., Bhagwat, Y. and Zhang, X. (2015).** Regaining “Lost” Customers: The Predictive Power of First-Lifetime Behavior, the Reason for Defection, and the Nature of the Win-Back Offer. *Journal of Marketing*, 79(4), pp.34-55.
- Kumar, V., Reinartz, W.J. (2006).** *Customer Relationship Management: A Databased Approach*, John Wiley and Sons Inc., New York.
- Kim, K., Jun, C. and Lee, J. (2014).** Improved churn prediction in telecommunication industry by analyzing a large network. *Expert Systems with Applications*, 41(15), pp.6575-6584.
- Lazarov, V., Capota, M. (2007).** Churn Prediction, *Bus. Anal. Course TUM Comput. Sci.*
- Lindmark, S., Andersson, E., Bohlin, E. and Johansson, M. (2006).** Innovation system dynamics in the Swedish telecom sector. *info*, 8(4), pp.49-66.
- Lopez, J.P.M., Redondo, Y.P. and Oliván, F.J.S. (2006).** The impact of customer relationship characteristics on customer switching behavior: differences between switchers and stayers, *Managing Service Quality*, Vol. 16, pp. 556-74.
- Lorena, A. C., & Carvalho, A. C. P. L. F. d. (2007).** Protein cellular localization prediction with support vector machines and decision trees, *Computers in Biology and Medicine*, 37, 115–125.
- Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E. and Kaushansky, H. (2000).** Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry, *IEEE Transactions on Neural Networks*, Vol. 11 No. 3, pp. 690-6.
- Nie, G., Rowe, W., Zhang, L., Tian, Y. and Shi, Y. (2011).** Credit card churn forecasting by logistic regression and decision tree, *Expert Systems with Applications*, 38(12), pp.15273-15285.
- Nie, G., Zhang, L., Li, X., & Shi, Y. (2006).** The analysis on the customers churn of charge email based on data mining – Take one internet company for example. In *Sixth IEEE international conference on data mining*, Hong Kong, China (pp. 843–847).
- Novak, T.P. and Hoffman, D.L. (2000).** Advertising and pricing models for the web, in Hurley, D., Kahin, B. and Varian, H. (Eds), *The Economics of Digital Information and Intellectual Property*, MIT Press, Cambridge, MA.

- Noyan, F. and Şimşek, G.** (2014). The Antecedents of Customer Loyalty, *Procedia - Social and Behavioral Sciences*, 109, pp.1220-1224.
- Oktay, E., Ozer, H., Ozcomak, M. S.** (2006). Ataturk Universitesi ogrencilerinin cep telefonu abonelik turunu tercih etmeleriyle ilgili faktorlerin tespiti”, *EKEV Akademi Dergisi*, Vol. 10 (27), pp. 275-296.
- Oliver, R.L.** (1999). Whence customer loyalty?, *Journal of Marketing*, Vol. 63, pp. 33-44.
- Ploetner, O.** (2004). NEGBI – introducing new systems in the telecom market, *Journal of Business & Industrial Marketing*, 19(5), pp.344-350.
- Reinartz, W.J. and Kumar, V.** (2002). The mismanagement of customer loyalty, *Harvard Business Review*, Vol. 80, pp. 86-94.
- Richeldi, M., & Perrucci, A.** (2002). Churn analysis case study, *Telecom Italia Lab Report*, Torino, Italy.
- SAS International** (2001). Predicting churn, SAS Institute, Cary, NC.
- Thomas, J.S., Blattberg, R.C. and Fox, E.J.** (2004). Recapturing lost customers, *Journal of Marketing Research*, Vol. 41, pp. 31-45.
- Turban, E., Leidner, D., McLean, E. and Wetherbe, J.** (2008). Information technology for management, *Transforming Organizations in the Digital Economy*, 6th ed., John Wiley and Sons, Hoboken, NJ.
- Vafeiadis, T., Diamantaras, K., Sarigiannidis, G. and Chatzisavvas, K.** (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, pp.1-9.
- Wei, C. -P., Chiu, I. -T.** (2002). Turning telecommunications call details to churn prediction: A data mining approach, *Expert Systems with Applications*, 23(2), 103–112.
- W.-H. Au, K.C. Chan, X. Yao.** (2003). A novel evolutionary data mining algorithm with applications to churn prediction, *IEEE Trans. Evol. Comput.* 7 (6), pp. 532–545.
- Wu, D.** (2009). Supplier selection: A hybrid model using DEA, decision tree and neural network., *Expert Systems with Applications*, 36, 9105–9112.
- Zakaria, I., Rahman, B., Othman, A., Yunus, N., Dzulkipli, M. and Osman., M.** (2014). The Relationship between Loyalty Program, Customer Satisfaction and Customer Loyalty in Retail Industry: A Case Study, *Procedia - Social and Behavioral Sciences*, 129, pp.23-30.

Zeithaml V.A., Berry L.L., Parasuraman, A. (1996). The behavioral consequences of service quality, *Journal of Marketing*, Vol. 60, pp. 31-46.

Zorn, S., Jarvis, W. and Bellman, S. (2010). Attitudinal perspectives for predicting churn, *Journal of Research in Interactive Marketing*, 4(2), pp.157-169.

Url-1 <<https://www.btk.gov.tr/uploads/pages/2017-q4.pdf>>, date retrieved 8.04.2018.

Url-2 <<http://www.ey.com>>, date retrieved 8.04.2018.



CURRICULUM VITAE

Name Surname : Cansu Alatas
Place and Date of Birth : Ankara
E-Mail : cnsalatas@gmail.com



EDUCATION :

- **B.Sc.** : 2014, Yildiz Technical University, Faculty of Naval Architecture and Maritime, Naval Architecture and Marine Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- **2017-Ongoing** Vodafone, Turkey- Product Management, Consumer Postpaid
- **2016-2017** PricewaterhouseCoopers, Turkey- Risk Assurance Services Consultant
- **2017** 5th Ace of M.I.C.E. – the Best Launch Event- 1st Place-as Mercedes Benz Turk
- **2016** 9th ITU UCARI Case Analysis Competition- 1st Place
- **2015** Global Social Venture Competition- Semifinalist in London, UK

