

**T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

BANKACILIKTA MÜŞTERİ TERK MODELİ

KÜBRA ŞEN

**YÜKSEK LİSANS TEZİ
MATEMATİK MÜHENDİSLİĞİ ANABİLİM DALI
SİSTEM ANALİZİ PROGRAMI**

**DANIŞMAN
YRD. DOÇ. DR. NİLGÜN GÜLER BAYAZIT**

İSTANBUL, 2014

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BANKACILIKTA MÜŞTERİ TERK MODELİ

Kübra Şen tarafından hazırlanan tez çalışması tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Haberleşme Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Yrd. Doç. Dr. Nilgün Güler Bayazıt

Yıldız Teknik Üniversitesi

Jüri Üyeleri

Yrd. Doç. Dr. Nilgün Güler Bayazıt

Yıldız Teknik Üniversitesi

Yıldız Teknik Üniversitesi

Yıldız Teknik Üniversitesi

ÖNSÖZ

Tez çalışmam boyunca her konuda yol göstericiliği ve desteği için danışmanım Yrd. Doç. Dr. Nilgün Güler Bayazıt'a, ayrıca maddi manevi desteklerini esirgemeyen aileme ve arkadaşlarıma sonsuz teşekkürlerimi sunarım.

Temmuz, 2014

Kübra ŞEN

İÇİNDEKİLER

	Sayfa
SİMGE LİSTESİ.....	vi
KISALTMA LİSTESİ.....	vii
ŞEKİL LİSTESİ.....	viii
ÇİZELGE LİSTESİ	ix
ÖZET	x
ABSTRACT.....	xii
BÖLÜM 1	
GİRİŞ.....	1
1.1 Literatür Özeti	2
1.1.1 Yurt Dışı Önceki Çalışmalar	2
1.1.1.1 Öğretmenli Öğrenme Teknikleri ile Yapılmış Çalışmalar.....	2
1.1.1.2 Öğretmensiz Öğrenme Teknikleri ile Yapılmış Çalışmalar	3
1.1.1.3 Esnek Programlama(Soft Computing) Teknikleri ile Yapılmış Çalışmalar.....	4
1.1.1.4 Diğer Metotlar ile Yapılmış Çalışmalar	4
1.1.2 Yurt İçi Önceki Çalışmalar	5
1.2 Tezin Amacı	8
1.3 Hipotez	8
BÖLÜM 2	
MODELDE KULLANILAN TEKNİKLER	9
2.1 Karar Ağaçları	9
2.2 Lojistik Regresyon	10
2.3 Yapay Sinir Ağları	11
2.4 Rastgele Orman Algoritması	11
2.4.1 Rastgele Orman Tekniğinin Öne Çıkan Faydaları.....	12

BÖLÜM 3

VERİ KÜMESİNİN HAZIRLANMASI	14
3.1 Problemin Tanımı ve İşin Amacı.....	14
3.2 Veri Kümesi	17
3.2.1 Veri Kümesindeki Alanların Belirlenmesi ve Oluşturulması.....	17
3.2.2 Verilerin Alınacağı Zaman Aralığının Belirlenmesi	19
3.2.3 Veri Kümesinin Analiz Edilmesi ve Model İçin Veri Hazırlama.....	22
3.3 Model Girdilerinin Seçilmesi	25

BÖLÜM 4

MODELİN KURULMASI	27
4.1 Klasik Veri Madenciliği Yöntemleri İle Terk Modelleri.....	27
4.1.1 Karar Ağaçları	29
4.1.2 Yapay Sinir Ağları	31
4.1.3 Lojistik Regresyon	31
4.2 Klasik Veri Madenciliği Modellerinin Terk Modeli Karşılaştırması	32
4.3 Rastgele Orman Yöntemi İle Terk Modeli.....	34

BÖLÜM 5

SONUÇ VE ÖNERİLER	36
KAYNAKLAR.....	39

EK-A

VERİ KÜMESİ	42
A-1 Veri kümesindeki alanlar ve açıklamaları	42
A-2 Oluşturulan Trend Değişkenleri.....	47

EK-B

DEĞİŞKENLERİN KARAKTER ANALİZİ SONUÇLARI	49
B-1 Kategorik değişkenlerin karakter analizi sonuçları	49
B-2 Tarihsel değişkenlerin karakter analizi sonuçları.....	52

EK-C

KARAR AĞACI MODELİ AĞAÇ GÖRÜNTÜSÜ.....	54
C-1 KA-CHAID ağacı	54
C-2 KA-CART ağacı.....	55
ÖZGEÇMİŞ	56

SİMGE LİSTESİ

$\log$	Logaritmik fonksiyon
N	Ağaç birliğindeki ağaç sayısı
n	Her düğümde rastgele seçilen ağaç sayısı
p	Birlikteki ağaçlar arasındaki korelasyonun ortalama değeri
Θ_k	k.ağaç için seçilmiş rastgele vektör

KISALTMA LİSTESİ

CRM	Customer Relationship Management
ETL	Extract Transform Load
EG	Enterprise Guide
EM	Enterprise Miner
KA	Karar Ağaçları
Log-Reg	Lojistik Regresyon
OOTB	Out of bag
ROC	Receiver Operating Characteristic
RO	Rastgele Orman
SAS	Statistical Analysis System
YSA	Yapay Sinir Ağları
WEKA	Waikato Environment for Knowledge Analysis

ŞEKİL LİSTESİ

	Sayfa
Şekil 2.1 RO Genel mimairisi	12
Şekil 3.1 Veri Kümesinin oluşturulması.....	19
Şekil 3.2 Sabit zaman aralığı ve dinamik zaman aralığı [40]	22
Şekil 4.1 SAS EM Üzerinde Geliştirilmiş Terk Modeli Diagramı	28
Şekil 4.2 SAS EM Sample Nodu Sonucu	27
Şekil 4.3 SAS EM Model Karşılaştırma nod Sonuçları	32
Şekil 4.4 SAS EM ROC Eğrisi Karşılaştırma Sonuçları.....	33
Şekil 4.5 WEKA RO Model Sonuçları	34
Şekil 4.6 RO modeli ROC eğrisi.....	35

ÇİZELGE LİSTESİ

	Sayfa
Çizelge 1.1 Araştırma zaman çizelgesi	7
Çizelge 3.1 Davranış Segmenti ve Kriterleri	15
Çizelge 3.2 Davranış segmentleri ve kriterleri	15
Çizelge 3.3 Veri Grupları	17
Çizelge 3.4 Sabit zaman aralığı için bazı alanların korrelasyon analiz sonucu	21
Çizelge 3.5 Veri kümesindeki terk eden müşterilerin frekansı	23
Çizelge 3.6 Gün sayısı hesaplanmış değişkenler	24
Çizelge 3.7 Gün sayısı değişkenlerinin Terk_f alanı ile korrelasyon analizi sonuçları	25
Çizelge 3.8 Model girdi listesi	26
Çizelge 4.1 KA-CART modelinin karmaşıklık matrisi	29
Çizelge 4.2 KA-CART Modeli değişkenlerin önem değerleri	30
Çizelge 4.3 KA-CHAID Modeli karmaşıklık matrisi	30
Çizelge 4.4 KA-CHAID Modeli değişkenlerin önem değerleri	31
Çizelge 4.5 YSA modelinin karmaşıklık matrisi	31
Çizelge 4.6 Lojistik regresyon modelinin karmaşıklık matrisi	32
Çizelge 4.7 Modellerin AUROC ve Doğruluk Oranı Karşılaştırması	33
Çizelge 4.8 RO karmaşıklık matrisi	35

BANKACILIKTA MÜŞTERİ TERK MODELİ

Kübra ŞEN

Matematik Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Yard. Doç. Dr. Nilgün GÜLER BAYAZIT

Değişik sektörlerde yapılan araştırmalar, bir şirket için mevcut müşteriyi elde tutabilmenin, yeni müşteri kazanımından daha maliyetli olduğunu göstermiştir. Dolayısıyla şirketlerin iyi bir müşteri elde tutma stratejilerinin olması ve bunu pazarlama/müşteri ilişkileri yönetimi ile uygulamaya geçirmesi büyük önem taşımaktadır. İyi bir elde tutma stratejisine en iyi girdi ise, terk etme potansiyeli taşıyan müşterileri belli bir güven aralığında modelleyebilmektir. Bu müşteriler belirlendikten sonra gerekli önlem ve aksiyonlar alınabilir.

Bu tez, Türkiye’de bankacılık sektörü için bir terk tahmini modeli geliştirmeyi amaçlamaktadır. Bankacılık sektöründe müşteri banka ilişkisinin süresi bir kontrata dayalı olmadığından, bu modeli bankacılık için geliştirmek diğer sektörlerle oranla daha zorlu bir süreçtir. Model için öncelikle müşteri ham verilerinin kullanışlı ve anlamlı bir hale dönüştürülerek veri ambarına alınması çalışması yapılmıştır. Daha sonra hazırlanan bu veri kümesi üzerinde, veri madenciliği teknikleri kullanılarak bir terk tahmin modeli oluşturulmuştur. Klasik tekniklerin yanı sıra son dönemde öne çıkan rastgele orman tekniği de kullanılarak sonuçlar tahmin başarısı açısından karşılaştırılmıştır.

Son olarak terk olasılığı tahmin edilmiş müşterilerin elde tutulabilmesi için, terke neden olan faktörler ışığında gerekli önlemlerin alınması ve stratejilerin geliştirilmesine yönelik önerilerde bulunulmuştur.

Anahtar Kelimeler: Müşteri terk tahmini, müşteriye elde tutma, CRM, veri madenciliği, karar ağaçları, yapay sinir ağları, lojistik regresyon, rastgele orman

CUSTOMER CHURN MODELLING IN BANKING

Kübra ŞEN

Department of Mathematical Engineering

MSc. Thesis

Adviser: Assistant Professor Dr. Nilgün GÜLER BAYAZIT

Studies across a number of industries have revealed that the cost of keeping an existing customer is more than the cost of acquiring a new one. Therefore, it is essential for a company to have a good retention strategy and apply this strategy through marketing and CRM tools. A crucial input to a good retention strategy is modeling the percentage of customers who are more prone to churn. Once these type of customers determined, then the necessary steps can be taken to prevent the churn .

This thesis is aiming to provide a churn prediction model for banking sector in Turkey. It proves a bit more challenging to provide a model for banking sector as there are no contractual agreements between a customer and a bank regarding the duration of services. During the implementation of the model, we first converted the raw data into usable and meaningful form so it could be used for our analysis. Later, using data mining techniques and our model, we build a “churn prediction model” for each customer in this data set. Besides the classical datamining techniques, we also use the recently popular “Random Forest” technique in our prediction, and compared the results in terms of prediction ability.

Lastly, for the purpose of improving retention performance, we propose some measures and make suggestions regarding the factors that are contributing more to churn according to our model.

Keywords: Customer churn prediction, customer retention, CRM, data mining, decision trees, neural networks, regression, random forests

BÖLÜM 1

GİRİŞ

Müşteriyi elde tutma yöntemleri ile müşteri terkinin engellenmesi CRM'in temel konularından biridir. Çeşitli araştırmalar göstermiştir ki yeni müşteri kazanmak her zaman eldeki müşteriyi tutmaktan çok daha maliyetlidir[1]. Bununla birlikte bir şirketteki müşteri devamlılığındaki %5 oranındaki artış, müşterinin değerinde %25 ila %95 oran aralığında bir artış sağlamaktadır [2].

Çapraz satış, fazladan satış aktiviteleri, ödüllendirmeler ve müşteri segmentasyon programları müşteri arttırmaya yönelik olsa da, her zaman bu eforların odağında aslında müşteriyi kaybetme korkusu vardır. Müşteri terki yönetiminde odaklanılan nokta, bir firma için hangi müşterinin kaybedilme olasılığının yüksek olduğunun bulunması ve bu müşterilerin elde tutulması için yapılacak çalışmanın maliyetinin analiz edilmesidir[1].

Müşterilerin şirketi terk edebileceği ve belirli bir geri kazanım eforu gösterilmediğinde geri dönmeyeceği her sektör için müşteri terkinin yüksek oranda olması kritik önem arz etmektedir. Bankacılık sektörü, telekomünikasyon sektörü, sigorta sektörü, yatırımcılık sektörü, internet servis sağlayıcıları ve gazetecilik/yayıncılık sektörüne kadar bir çok sektör müşteri terk probleminin yönetilmesi gereken sektörler arasındadır.

İstemli ve istemsiz olmak üzere başlıca iki çeşit müşteri terki vardır. İstemsiz terk çeşitinde şirket müşteri ile ilişkisini kendisi bitirmeye karar verir. İstemli terk çeşitinde ise müşteri şirket ile ilişkisini bitirmeye karar verir. Hadden vd. [3] te istemli terk çeşitini de planlanmış ve tesadüfi olarak da ikiye ayırmıştır. Planlanmış istemli terkte, müşteri şirketten memnun değildir ya da daha iyi bir teklif almıştır. Tesadüfi istemli

terkte ise müşterinin artık şirketin sağladığı ürüne ihtiyacı kalmamıştır ya da şirketin hizmet vermediği bir bölgeye taşınmıştır.

1.1 Literatür Özeti

1.1.1 Yurt Dışı Önceki Çalışmalar

Terk edecek müşterilerin tahmin edilmesine yönelik yapılan çalışmalarda hem öğretmenli hem de öğretmensiz öğrenme teknikleri sıklıkla kullanılmıştır. Öğretmenli öğrenme tekniklerinde müşterilerin terk eden müşteri ve terk etmeyen müşteri olarak etiketlendiği bir veri kümesi kullanılarak bir model kurulmakta ve bu model daha sonra etiketlenmemiş müşterilerin veri kümesi üzerinde kullanılarak terk edecek müşterilerin tahmin edilmesi sağlanmaktadır.

Geeta, V. ve Shashi, M., [4] de öğretmenli öğrenme tekniklerinden müşteri terki tahmini için en sık kullanılan tekniklerden bazılarını ve bu teknikler ile yapılan çalışmaları aşağıdaki şekilde gruplamıştır.

1.1.1.1 Öğretmenli Öğrenme Teknikleri ile Yapılmış Çalışmalar

- **Karar Ağaçları**

Kitayama, M. vd. [5] de verilen çalışmalarında karar ağacına dayalı bir yaklaşım kullanarak müşteri profilleri oluşturmuşlar ve daha sonra bu segmentler için yine karar ağacı uygulayarak müşterilerin başka bir şirkete geçmelerini engellemek için alınması gereken önlemleri belirlemişlerdir. Hwang, H. vd. [6] de karar ağacı, yapay sinir ağları ve lojistik regresyon içeren araştırmalarında karar ağaçlarının müşteri segmentasyonunda çok daha doğru bir sonuç vermiş olduğunu göstermiş ; ancak her durumda en iyi seçimin her zaman karar ağaçları olmayabileceğini de belirtmişlerdir. Ng, K. ve Liu, H. [7] te müşteriyi elde tutma amacı taşıyan çalışmalarında özellik seçiminin çok doğru olması gerektiğini belirterek, C4.5 karar ağacı ile sınıflandırma yapmış ve iyi sonuç verdiğini kanıtlamışlardır. Datta, P. vd. [8] te CHAMP (Churn Analysis Modelling and Prediction) adını verdiği modellerinde telekom sektöründe terk tahminlemesi için karar ağaçlarını kullanmıştır.

- **Regresyon**

Kim, H. ve Yoon, C. [9] de müşteri memnuniyetini müşteriye elde tutma özellikleri ile bağdaştırmak için lojistik regresyon kullanmışlardır. Müşteri elde tutma olasılığının lojistik fonksiyonuyla çıkarılabileceğini iddia etmişlerdir. Ayrıca telekomünikasyon sektöründe terk edecek müşterilerin tespit edilmesi için ayrık seçim teorisine dayalı bir lojistik model geliştirmişlerdir. Mutanen T. [10] de bireysel bankacılıkta lojistik regresyon uygulayarak terk tahmini yapmış ve modelin etkisini değerlendirmek için lift curve kullanmıştır. Hadden, J. [11] da müşteri şikâyet ve çözüm verilerini kullanarak terk tahminlemesi üzerinde çalışmışlardır. Bu çalışma genellikle müşteri demografik bilgileri üzerine yoğunlaşan önceki çalışmalardan farklı bir çalışma olmuştur. Bu çalışmada, lineer regresyon, regresyon ağaçları ve yapay sinir ağları gibi çeşitli terk modelleri kullanılmış olup terk tanımlamasında Bayesian çerçeveli yapay sinir ağlarının, terk tahminlemesinde ise regresyon ağacının en iyi performansı gösterdiği kanıtlanmıştır.

- **Bayes Teoremi**

Nath, S. [12] de Oracle 9i veri madenciliği ve Naïve Bayes algoritması kullanarak müşteri terk problemi üzerinde çalışmıştır, model %68 doğruluk göstermiştir. Hadden, J. [13] de müşteri terkinin önceden tahmin etmek için hatalı sınıflandırma oranını minimumda tutan bir müşteri profillemesi methodu geliştirmiştir. Müşteri terki tahmini için yapay sinir ağları, regresyon ağaçları ve lineer regresyon modellerini kullanmış, terk tahmininde en başarılı sonuç veren modelin Bayesian yaklaşımlı yapay sinir ağları olduğunu göstermiştir.

1.1.1.2 Öğretmensiz Öğrenme Teknikleri ile Yapılmış Çalışmalar

Öğretmensiz öğrenme teknikleri genellikle bir hibrid modelin parçası olarak ve en çok karar ağaçları ile bağlantılı olarak kullanılmaktadır [4].

Bose I., ve Chen X., [14] de müşteri terk tahmini için, kümeleme kullanılan hibrid modellerin kullanılmayanlara kıyasla daha iyi sonuç verdiğini göstermiştir. Hung S.Y. vd. [15] de bir değişkene göre müşterileri kümelemiş ve her küme için bir karar ağacı tasarlamışlardır. Hangi kümenin daha doğru bir tahminleme geliştireceğini bulmak için

karar ağacını aynı test verisi üzerinde kullanmışlardır. Chu B. H., vd. [16] de müşteriye elde tutma problemini geliştirmek için bir hibrid yapı sunmuşlardır. Modelin amacı terk olasılığını tahminlemek olduğu gibi aynı zamanda müşteriye elde tutma stratejileri de sunmaktır. Terk modeli karar ağacına dayalı bir sınıflandırma tekniği ve GHSOM (Growing Hierarchical Self Organising Map) e dayalı bir kümeleme modeli kullanılarak oluşturulmuştur.

1.1.1.3 Esnek Programlama(Soft Computing) Teknikleri ile Yapılmış Çalışmalar

Bulanık mantık, yapay sinir ağları ve genetik algoritmalar gibi bazı yöntemler de müşteri terk tahmini çalışmalarında kullanılmıştır [12].

Au, W. vd. [17] de müşteri terk tahminlemesi üzerine yaptıkları çalışmada yapay sinir ağlarının karar ağaçlarından ve lojistik regresyondan daha iyi sonuç verdiğini göstermişlerdir. Rygielski, C. vd. [18] daki çalışmalarında yapay sinir ağlarının çok daha güçlü ve kesin bir tahminleme yapabileceğini; ancak karar ağaçları ile karşılaştırıldığında yetersiz sonuçlar bulmakta ve aşırı öğrenme gibi risklerinin olduğunu göstermişlerdir. Shin Yuan H. vd. ise [19] de müşteri demografik bilgileri, ödeme bilgileri ve kontrat/servis durumları, arama kayıtları ve servis değişim loglarını içeren bir modelde hem yapay sinir ağlarının hem de karar ağaçlarının oldukça kesin sonuç verebildiğini göstermişlerdir. Džulijana P., ve Bojana Dalbelo B., [20] de Fuzzy C-Means tekniğine dayalı 5000 müşteri içeren bir bireysel bankacılık verisi üzerinde uygulanan bir model sunmuşlardır.

1.1.1.4 Diğer Metotlar ile Yapılmış Çalışmalar

Chiang, D. vd. [21] da çalışmalarında potansiyel terk edecek müşterilerin belirlenebilmesi için Goal- Oriented Sequential Pattern olarak isimlendirilmiş yeni bir algoritma sunmuşlardır. Bu çalışmada değişkenler arasındaki ilişkileri belirleyen ilişkili kurallar kullanılmıştır. Madden, G. ve Savage, G. [22] da Avusturalya İnternet Servis Sağlayıcı endüstrisi bireysel müşterileri için bir terk modeli sunmuştur. Modelde ekonomik veriler, servis kullanım verileri, servis seçim ölçüt verileri ve müşteri demografik verileri kullanılmıştır. Tao, Y. H. ve Yeh, C. R. [23] de müşteri davranışlarını anlama ve buna bağlı olarak müşteriye elde tutma yöntemleri üzerine çalışmışlardır.

USC(Usage Segment Code) ve NRE (Net Revenue Equation) adında iki uygulama kullanılmıştır. Bu uygulamalar terk tahmininden öte müşterinin ikna edilmesi için gerekli stratejilerin geliştirilmesine yardım etmek için kullanılmıştır.

1.1.2 Yurt İçi Önceki Çalışmalar

Özmen [24] te karar ağaçları ve regresyon modelleme yöntemlerini kullanarak bir mobil iletişim şirketi müşterilerinin her birine bir puan atamış ve bu puanlar üzerinden müşterinin 6 ay sonra ayrılıp ayrılmayacağını tahminlemiştir.

Tosun [25] te özel bir bankaya ait müşteri verilerini kullanarak kredi kartı müşterilerinin çeşitli niteliklerdeki bilgilerini inceleyerek, kaybedilmiş bir müşterinin profili veri madenciliği yöntemleriyle ortaya çıkarmaya çalışmıştır.

Kişioğlu [26] da Bayes ağları methodunu kullanarak çıkan analiz sonuçlarına göre üç farklı senaryo oluşturarak iptali azaltmaya yönelik önerilerde bulunmuştur.

Bilgen [27] de bir banka müşterilerinin e-bankacılık sistemi kullanımı verileri üzerinde FAHP (Fuzzy Analytic Hierarchy Process – Analitik Hiyerarşi Yöntemi) metotunu kullanarak müşterilerin tatminliğini ölçerek davranışlarını tahminlemiştir.

Koçtürk [28] de 3 farklı karar ağacı algoritması kullanarak bireysel emeklilik sahibi müşterilerin sigorta şirketinden ayrılarak rakip şirketlerin müşterisi olmayı tercih eden davranışlarını ortaya koyan modeller geliştirilmiştir.

Güllüoğlu [29] da bir bankanın 6 aylık müşteri verilerini kullanarak karınca koloni algoritması (ant colony algorithm) kullanarak müşterileri sınıflandırmıştır; ayrıca Naive Bayes metodu ile de sınıflandırma yapmıştır.

Arifoğlu [30] da bir telekom firması müşteri demografik ve arama verileri ile veri madenciliği tekniklerini kullanarak terk tahmini yapmıştır. Karar ağaçları, Naive Bayes yöntemi, destek vektör makineleri (SVM), olasılı yapay sinir ağları (Probabilistic NN) ile karşılaştırmalarını yapmış ve en iyi sonuç olarak ANFIS Fuzzy C-means algoritmasının sonuç verdiğini göstermiştir.

Arıtürk [31] de herbir müşteri için farklı zaman dilimlerine ait birden fazla sayıda gözlemin kullanıldığı boylamsal veri seti üzerinde yapılan ‘gelecek dönem müşteri

kaybı' modellemesinin sentetik üst örneklemenin uygulandığı ve uygulanmadığı geleneksel kesitsel veri kullanımına kıyasla tahmin performansını artıracakını ve yöneticilerin ilave çıkarımlarda bulunmasına imkân sağlayacağını göstermiştir. Çizelge 1.1'de yukarıda bahsedilen çalışmaların yıllara göre listesi verilmiştir.

Çizelge 4.1 Araştırma zaman çizelgesi

Yıl	Öğretmenli Öğrenme Teknikleri			Enek Programlama Teknikleri		Bulanık Mantık		Diğer Teknikler
	Karar Ağaçları	Regesyon	Bayes Teoremi	Öğretmensiz Öğrenme Teknikleri	ANN			
1993		Rust ve Zahorik						
1999								Madden ve Savage (Binomial probit modeli)
2000	Mozar vd.	Mozar vd.						
2001	Ng ve Lin	Mihelis vd.						
	Datta vd.							
2002	Kitayama vd.				Rygielski			
2003			Nath & Behara					Taove Yeh (USC ve NRE)
2004	Hwang vd.	Hwang vd.	Baesens vd.			Shin ve Sohn		
	Liao ve Chan	Kim ve Yoon						
2006		Mutanen		Thomassey and Fiordaliso (Hibrid Model ve Karar Ağacı ile)	Au vd.	Roy vd.		Prinzis and Van Den Poel (MTD)
		Juayin Qi		Hung vd. (Hibrid Model, Karar Ağacı ile)	Shin vd.			Lee vd.(SVM)
		Hadden vd.			Roy vd.			
					Hadden vd.			
2007				Chu vd. (Hibrid Model, SOM ve Karar Ağacı ile)				
2009			Kişioğlu	Bose Indrani ve Chan Xi			Dzulijana Popovic ve Bojana Dabelo Basic	Bilgen
2010	Koçtürk							
2011	Artürk	Artürk				Artıoğlu		Güllüoğlu

1.2 Tezin Amacı

Tezin amacı bankacılık sektöründe etkili ve verimli bir müşteri terk tahmin modeli kurmaktır. Türkiye’deki özel bir banka için öncelikle bir terk tanımı belirleyerek, gerçek müşteri verileri analiz edilerek terk tahmininde kullanılmak üzere yeni bir veri kümesi oluşturulmuştur. Bu veri kümesi üzerinde veri madenciliği teknikleri kullanılarak terke neden olan faktörler ışığında bankayı terk edecek müşterilerin önceden tahmin edilmesini sağlayacak model kurulmuştur.

1.3 Hipotez

Bu tezde ilk olarak bankanın en önemli iki segmentindeki müşteriler için terk tanımı yapılmıştır. Daha sonra bu terk tanımına yönelik oluşturulacak modelde kullanılmak üzere yeni bir veri kümesi hazırlanmış ve anlamlı, modelde kullanılabilecek değişkenler seçilmiştir. Bu veriler her müşteri için sabit bir zaman aralığında alındığında modelin çok etkili olmayacağı ve terk tahmini için yeterli olmadığı gösterilmiştir. Böylece her müşteri için kullanılacak verilerin kendi terk tarihinden önceki 6 aylık döneme ait olması gerektiği belirlenmiş olup her müşteri için dinamik bir zaman aralığında veriler oluşturulmuştur.

SAS uygulamasının Enterprise Guide 6.1 [32] aracı kullanılarak hazırlanan veri üzerinde gerekli önhazırlık çalışmaları yapılarak değişken seçimi, veri temizleme ve mevcut değişkenlerden yeni değişken oluşturulması gerçekleştirilmiştir. Veri model için hazırlandıktan sonra SAS uygulamasının Enterprise Miner 6.1 [32] aracı üzerinde klasik veri madenciliği yöntemlerinden karar ağaçları, yapay sinir ağları ve regresyon teknikleri ile geliştirilmiş, modellerin terk tahmini açısından sonuçları karşılaştırılmıştır. Bununla birlikte WEKA 3.6 [33] uygulaması kullanılarak son zamanlarda bankacılık sektöründe terk tahmini modellerinde öne çıkan rastgele orman yöntemi ile de aynı veri kümesi için başka bir terk tahmin modeli kurularak diğer üç model açısından karşılaştırılmaları yapılmıştır. Karşılaştırmalar sonucu terk tahmininde en iyi sonucu veren tekniğin rastgele orman metodu olduğu görülmüştür.

Modellerin sonucunda elde edilen bilgiler ışığında da müşteriyi elde tutmak için yapılması gerekenler hakkında öneriler getirilmiştir.

BÖLÜM 2

MODELDE KULLANILAN TEKNİKLER

Müşteri ilişkileri yönetimi (CRM) ve özellikle müşteri terki tahmini konusu son yıllarda fark edilir bir büyüme göstermiştir. Müşteri terki tahmini modelleri müşterilerin geçmiş davranışlarını inceleyerek hangi müşterilerin terk etmek üzere olduğunun belirlenmesine dayanır. En etkili ve geçerli terk tahmin modeli araştırılırken pek çok sayıda veri madenciliği tekniği kullanılmış olup, lojistik regresyon, karar ağaçları, yapay sinir ağları, destek vektör makineleri ve rastgele orman bu konuda öne çıkan tekniklerdir [34].

2.1 Karar Ağaçları

Karar ağaçları karar çıkarma ve görsel olarak ifade edilebilmesi gibi özellikleri nedeniyle en popüler veri madenciliği araçlarından biridir. Karar ağaçlarının başlangıcı, 1963'te Morgan ve Sonquist tarafından geliştirilen ağaç temeli AID (Automatic Interaction Detection) isimli bir modele dayanmaktadır. Bu modelin amacı bağımsız değişkenler arasındaki ilişkileri etkili bir şekilde ifade etmektir. Daha sonra 1975'te Hartigan AID modelini, veriyi alt kümelere bölecek şekilde istatistiksel olarak daha geçerli bir model haline dönüştürmüştür. Bu modele CHAID (Chi-Square Automatic Interaction Detection) denmektedir ve veriyi bölmek için ki-kare (chi-square) testi kullanmaktadır. 1984'te Breiman vd. tarafından geliştirilen CART isimli ağaç ile karar ağaçları daha da popüler hale gelmiştir. CHAID'den farklı olarak CART, büyük ölçülü ağaçlar geliştirir ve daha sonra dâhili bir çapraz geçerlilik yöntemi kullanarak aşırı öğrenme olasılığını engellemek için bazı dalları budar.

Karar ağaçlarındaki en önemli araştırma konusu en iyi bölme kuralının bulunabilmesidir. Bölme kuralı, tahmin değişkeninin hangi değer üzerinden bölünmesi gerektiğini belirleyen metottur. Bir ağaç geliştirmek için her düğümde veriyi iki alt düğüme bölmek için bir test ölçütünün uygulanması gereklidir. En çok kullanılan bölme kuralları ise Gini katsayısı kullanarak bölme yapılması, entropi değeri kullanılarak bölme yapılması ve ki-kare testi ile bölme işlemi yapılmasıdır [1].

2.2 Lojistik Regresyon

Regresyon da karar ağaçları gibi veri madenciliğinin en popüler tekniklerinden biridir. Lojistik regresyonda sadece bir bağımlı değişken olabilir ve bu bağımlı değişkeni lojistik bir değere çevirdikten sonra maksimum olabilirlik kestirimi uygulanır. Lojistik regresyonda, normal regresyondan farklı olarak bağımlı değişken ikilidir; bağımlı değişken q olasılığı ile 1, ya da 1- q olasılığı ile 0 değerini alabilir.

Normal regresyon aşağıdaki denklem ile gösterilebilir.

$$X = (x_1, x_2, \dots, x_n) \text{ iken}$$

$$P(x) = \alpha + \beta x \quad (2.1)$$

P(x) değerinin, X ile monoton şekilde değişerek 0 ile 1 arasında kısıtlanması için, lineer fonksiyon dışında bir lojistik fonksiyon kullanılmalıdır [10].

$$P(X) = \frac{e^{-(\alpha + \beta x)}}{1 + e^{-(\alpha + \beta x)}} \quad (2.2)$$

$$Q(X) = 1 - P(X) = \frac{1}{1 + e^{-(\alpha + \beta x)}} \quad (2.3)$$

Bu durumda n girdili lojistik regresyon modelinde çıktı aşağıdaki formül ile gösterilebilir.

$$Q(X) = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (2.4)$$

2.3 Yapay Sinir Ağları

Yapay sinir ağları(YSA) veri madenciliğindeki en güçlü tekniklerden biridir. İnsan beyni örnek alınarak oluşturulmuş bir modeldir, YSA modelleri harici girdiler öğrenir ve genelleme yaparlar. Eğitim gözlemleri sunulduğunda YSA desenleri ve ilişkileri keşfedebilir, örneklerden öğrenerek önemli ilişkileri ortaya çıkaracak desenleri araştırır. YSA'nın en güçlü yanı yüksek oranda lineer olmayan ilişkileri önceden belirlenmiş çok az bir varsayım ile modelleyebilmesine dayanır.

2.4 Rastgele Orman Algoritması

Rastgele orman(RO) algoritması 2001'de Breiman tarafından geliştirilmiş olup, ağaç sınıflandırıcılarının bir kombinasyonu olarak tanımlanmaktadır. Sınıflandırıcılar üzerinde doğruluğu artırmaya yönelik yapılan çalışmalar, çok sayıda ağaç üretilip bu ağaçlar birleştirilerek en iyi sınıfın oylama ile bulunmasıyla sonuçlanmıştır [35].

Bu ağaç birliğinin geliştirilmesi için, birlikteki her ağacın gelişimini kontrol eden rastgele vektörler üretilir. Her ağacın gelişmesi için eğitim kümesinden rastgele bir seçimin yapıldığı bagging buna ilk örnek olarak verilebilir [35].

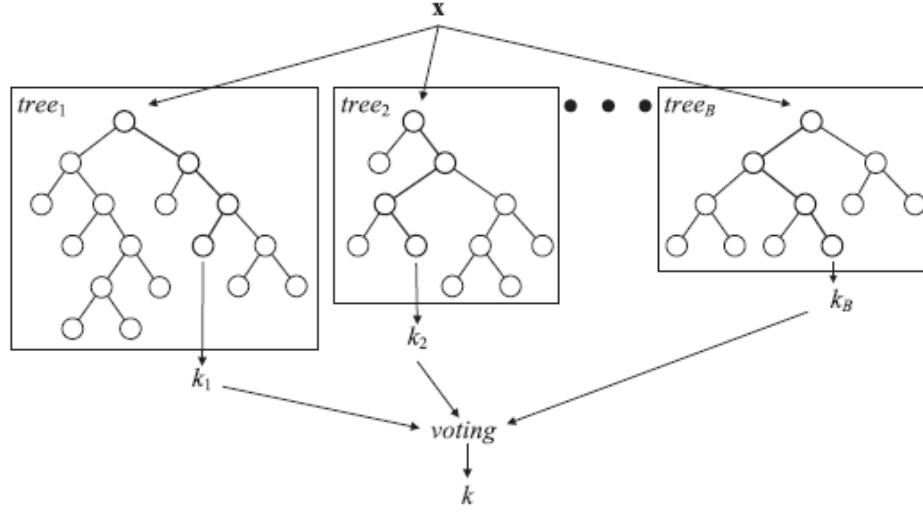
Diğer bir örnek ise rastgele bölme seçimidir ki; her düğümde bölme K tane en iyi bölünme arasından rastgele seçilir. Bu iki örnek dışında rastgele vektör seçimleri ile ilgili başka yöntemler üzerinde de çalışmalar da yapılmıştır.

Tüm bu yöntemlerin ortak unsuru, k. ağaç için geçmiş rastgele vektörlerden bağımsız ancak aynı dağılıma sahip bir Θ_k rastgele vektörünün üretilmesidir. Böylelikle eğitim kümesi ve Θ_k rastgele vektörü kullanılarak x 'in girdi olduğu $h(x, \Theta_k)$ ağaç sınıflandırıcısı geliştirilmiş olur. Çok sayıda ağaç geliştirildikten sonra en iyi sınıf için oylama yapılır. Bu prosedüre Rastgele Orman denir.

Önyüklem örnekleri uygulanarak Θ_k üretilirken her ağaç için eğitim veri setinin 2/3'ü kullanılır, geriye kalan 1/3'lük kısmı da OOB (out-of-bag) olarak adlandırılan kısımdır. OOB test veri seti olarak kullanılır ve hataları test eder. Birlikteki sınıflandırıcı ağaçlar arasındaki korelasyonun ortalama değerini gösteren p – faktörü düşük tutulursa test veri kümesindeki hata da büyük ölçüde azalacaktır.

Her ağaç olabilecek en büyük genişliğe kadar geliştirilir, budama yoktur.

Şekil 2.1 'de rastgele orman tekniğinin genel bir mimarisi verilmiştir [36].



Şekil 4.1 RO Genel mimairisi

B : RO'daki ağaç sayısı

k_1, k_2, k_B, k : sınıf etiketleri

Düşük öngörü ve düşük korrelasyon doğruluk için çok önemlidir. Düşük öngörü için ağaçlar maximum derinlikte geliştirilir. Düşük korrelasyon içinse rastgele seçim uygulanır [36]:

1. Her ağaç, eğitim kümesinin bir önyükleme örnekleme üzerinde geliştirilir.
2. Bir ağaç geliştirilirken, her düğümde, N değer içerisinde n tanesi rastgele seçilir.
3. Genellikle $n \leq N$ dir. $n = \lceil \log_2 N + 1 \rceil$ ya da $n = \sqrt{N}$ alınarak başlanması ve OOB veri kümesindeki hata değeri minimuma ulaşıncaya kadar artırılması ya da azaltılması tavsiye edilir.

2.4.1 Rastgele Orman Tekniğinin Öne Çıkan Faydaları

Rastgele orman'ın sınıflandırmada etkili olarak öne çıkmasındaki nedenlerden bazıları şunlardır [37]:

- Yüksek doğruluk oranı

- Büyük veri tabanlarında verimli olarak çalışabilmesi
- Değişken azaltmadan binlerce değişkenin üstesinden gelebilmesi
- Sınıflandırmada hangi değişkenlerin önemli olduğu konusunda tahmin verebilmesi
- Datanın büyük çoğunluğu eksik olsa da iyi sonuç verebilmesi
- Dengesiz veri kümelerinde hatayı dengeleyebilen metotlar üretebilmesi
- Ağaç sayısı ne kadar arttırılsa da aşırı öğrenme olmaması

BÖLÜM 3

VERİ KÜMESİNİN HAZIRLANMASI

Müşteri terkinin engellenmesi ve yönetilebilmesi için şirketlerin etkili ve geçerli bir terk modeli kurmaları önemlidir. Modelin uygulanmasından sonra ise terk edecek müşteriler için gerekli elde tutma stratejilerini geliştirmesi gerekmektedir. Terk analizi için ilk iş, terk etmiş olarak belirlenmiş müşterilerin tanımının yapılmasıdır. Terk tanımı yapıldıktan sonra modele en uygun veri kümesinin seçilmesi gerekmektedir. Veri kümesi seçimi çok önemli bir adım olup işin neredeyse %60 - %70 civarını oluşturmaktadır. Veri kümesi değişkenlerinin seçimi gibi veri kümesinin alınacağı zaman aralığının seçilmesi de kritik önem arz etmektedir. Bazı çalışmalarda dinamik zaman aralıkları (her müşterinin terk etmeden önceki 3 ayı içindeki hareketlerinin alınması vs.) belirlenebildiği gibi, tüm müşteriler için aynı zaman aralığının dikkate alındığı araştırmalar da mevcuttur.

Bankacılık sektöründe müşteri ile banka arasında kontrata dayalı bir ilişki olmadığından bankayı terk edecek müşterilerin belirlenmesi diğer sektörlerle göre daha zorlu bir çalışmadır [11]. Çoğu finansal şirkette müşteri terk tanımı yapılırken müşterinin ürün aktifliği kullanılır ve yine şirket tarafından belirlenmiş bir aktiflik eşiği belirlenerek bu eşiğin altına düşen müşteriler terk olarak belirlenir [38].

3.1 Problemin Tanımı ve İşin Amacı

Bu çalışmada, özel bir banka için bir terk tanımı belirlenerek, terke neden olan faktörlerin ışığında terk etmesi muhtemel müşterilerin tahmin edilmesini sağlayacak bir model kurulmuştur.

Terk tanımı yapılırken, müşterilerin değerlerine ve davranışlarına göre segmente edilmiş olmalarından faydalanılacaktır. Müşterilerin firmaya kazandırdıkları değere göre segmente edilmesine değer bazlı segmentasyon denilmektedir. Bankadaki değer segmentasyonu sonucu tüm bireysel müşteriler değerlerine göre segmente edilerek, her müşterinin dahil olduğu segment belirlenmiştir. Bir müşterinin değer segmenti bulunurken bir çok kriter göz önünde bulundurulmuştur.

Müşteri segmentleri ve değerlendirme kriterleri aşağıdaki tabloda belirtilmiştir.

Çizelge 4.2 Davranış Segmenti ve Kriterleri

DEĞER SEGMENT	KRİTER
ELMAS	($M^1 + K^2 \geq 100.000$ TL
ALTIN	$50.000 \text{ TL} \leq (M + K) < 100.000 \text{ TL}$
GÜMÜŞ	$15.000 \text{ TL} \leq (M + K) < 50.000 \text{ TL}$
BRONZ	$(M + K) < 15.000 \text{ TL}$
GİRİŞİMCİ	Ticari Kriterlere Uyum veya Nakdi Kredi Limiti $< 100.000 \text{ TL}$
ORTAK HESAP	Ortak Hesaplar

Değer segmentasyonu gibi müşterilerin davranışsal özelliklerine göre de segmentasyon yapılmaktadır.

Çizelge 4.3 Davranış segmentleri ve kriterleri

DAVRANIŞ SEGMENTİ	KRİTER
Kayan yıldızlar	Mevduat trendi düşüşte olan müşterilerdir.
Hareketsizler	Kanal kullanım adetleri çok az olan ve EFT/Havale tutarları çok az olan müşterilerdir.
Yıldızlar	Gelişmiş ürün sahipliği ³ olan veya orta seviye ürün ¹ kullanıp ürün sayısı 2 den fazla olan ve sadece mevduatçı olmayan müşterilerdir.

¹ Son 1 yıllık mevduat ortalaması

² Son 1 yıllık kredi ortalaması

³ Mevduat ve kredi kartının yanında kredi, kiralık kasa, talimat, sigortadan en az birinin olmasıdır.

Mütevaziler	Gelişmiş veya orta seviye ürün kullanımına sahip olmayan, çoğunlukla mevduat ürünlerini kullanan müşterilerdir.
İnaktifler	Ürün aktifliği olmayan(mevduat ürünleri için bakiyesi olmayan, diğer ürünler için aktifliği olmayan), kanal kullanımı olmayan, kanal kullanım trendi olmayan ve kanal kullanım tutarı olmayan müşterilerdir.

Değer ve davranış segmentasyonları bazında müşterilerin ürün sahiplikleri, verimleri ve mevduat trendleri incelendiğinde farklı değer segmentindeki müşteriler için farklı terk tanımlarının yapılmasının daha doğru olacağı tespit edilmiştir. Değer segmentinde üst seviyede ve birçok ürün sahibi bir müşterinin terk tanımı ile sadece kredi kartı kullanan ve kartını kapattığında terk edecek bir müşterinin aynı terk tanımı ile belirlenmesinin doğru olmayacağı görülmüştür. Dolayısıyla her segmentteki müşterinin terk tanımı farklılık gösterecektir.

Bu bilgiler ışığında müşteri terk analizi yapılırken ilk olarak en değerli iki segment olan Elmas ve Altın segmentindeki müşteriler göz önünde bulundurulmuştur. Elmas ve Altın segmentindeki müşterilerin son aylardaki davranışları incelendiğinde bu müşterilerin farklı davranış segmentlerinde olabildikleri görülmüştür. Ayrıca davranışların tarihçesi incelendiğinde, müşterilerin davranış segmentlerinde aylık değişimlerin yaşandığı gözlemlenmiştir. Örneğin, inaktif davranış segmentine geçmiş Elmas ve Altın müşterilerden bazıları bir müddet inaktif segmentte kaldıktan sonra tekrar aktif segmentlerden birine geçebilmektedir.

Bu verilerden yola çıkılarak terk tanımı şu şekilde yapılmıştır: Elmas ve Altın değer segmentinde olup, inaktif davranış segmentine dahil olan yani ürün aktifliklerinin olmadığı, kanal kullanım trendi ve hesap bakiyelerinin sıfırlandığı duruma geçtikten sonra en az 6 ay boyunca inaktif segmentte kalan müşteriler terk eden müşteriler olarak belirlenmiştir.

Uygulamada kullanılan veri kümesi Elmas ve Altın müşterilerden seçilmiş ve yukarıdaki tanıma uyan müşteriler terk olarak işaretlenmiştir. Veri kümesi müşteri demografik bilgileri, kanal kullanım, ürün sahiplik ve bazı çevresel verilerden oluşturulmuştur. Çalışmada, ilgili veri kümesi üzerinde veri madenciliği algoritmaları kullanılarak terk

¹ Mevduat ve kredi kartının yanında ADK kullanımının olmasıdır

edecek müşterilerin tahmin edilmesi ve müşteri davranış segmentine göre elde tutma modeli geliştirilmesi amaçlanmaktadır.

3.2 Veri Kümesi

Terk edecek müşterilerin doğru tahmin edilmesi yani oluşturulacak terk modelinin başarısında veri kümesinin doğru seçilmesi çok büyük önem taşımaktadır. Bu nedenle verinin hazırlanması terk modelinin en önemli adımıdır [39].

3.2.1 Veri Kümesindeki Alanların Belirlenmesi ve Oluşturulması

Bu çalışmada, kullanılmak üzere, elmas ve altın değer segmentindeki müşterilerin bir çok verisi incelenmiş ve terki etkileyebilecek tüm veriler veri kümesine eklenmiştir. Veri kümesindeki işlem verileri, ürün sahipliği verileri, kanal kullanım verileri ve demografik veriler 9 ayrı tablodan alınarak birleştirilmiştir.

Veri grupları aşağıdaki Çizelge 3.3’de gösterilmiştir.

Çizelge 4.4 Veri Grupları

Veri Grubu	Alanlar
Demografik	Müşterinin yaşı, cinsiyeti, eğitimi, mesleği, ünvanı, çalışma tipi, medeni hali, çocuk sayısı, bankaya yakınlık bilgisi, müşterilik yaşı vs.
İşlem /hareket verileri	Toplam gelen/giden eft/havale işlem trendi, kredi kartı harcama trendi, kredi ödeme bilgileri, fatura ödeme bilgileri vs.
Kanal kullanım	Şube, internet şubesi, ATM ve diğer kanalları kullanım trendi
Ürün sahipliği	Kredi kartı, mevduat, kredi, fatura vs. ürün sahiplik ve aktiflikleri, ürünleri açma/kapama tarihleri
Çevresel veriler	Bağlı olduğu şube bilgisi

Müşterilerin demografik verileri olarak müşterinin yaşı, cinsiyeti, eğitimi, mesleği, ünvanı, çalışma tipi, medeni hali, çocuk sayısı, bankaya yakınlık bilgisi, müşterilik yaşı, değer segmenti, davranış segmenti, geliri, uyruğu ve ev sahibi olup olmadığı bilgileri alınmıştır.

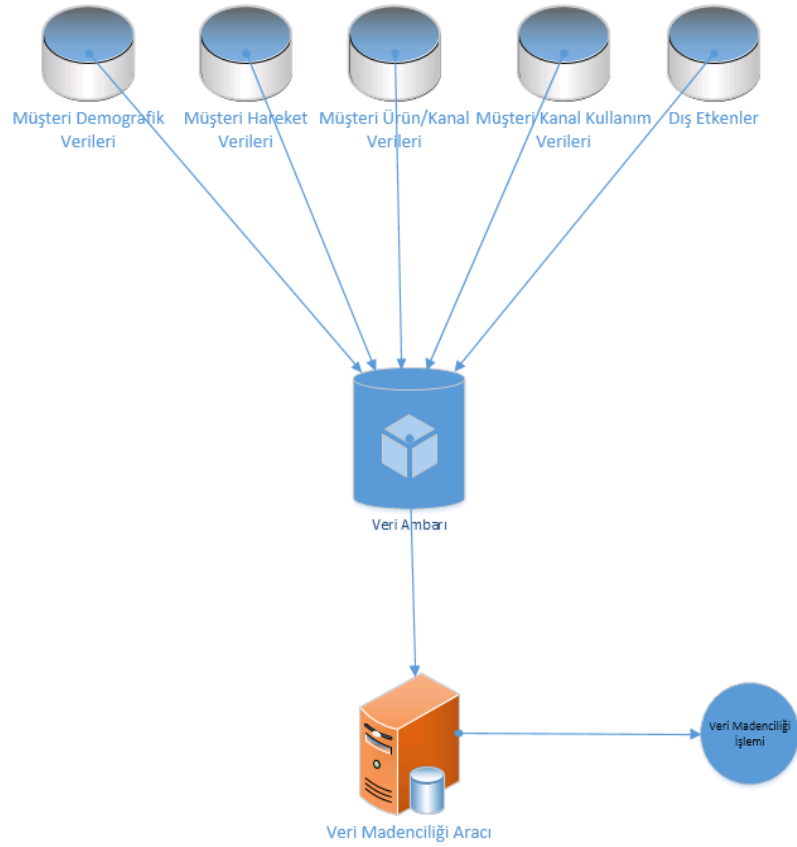
İşlem/hareket ve kanal kullanım bilgileri olarak müşterinin cari, katılım, yatırım ve toplam hesap bakiye ortalamaları, müşterinin bankaya verimi, maaş ödemesi

ortalaması, giden eft adedi, gelen eft adedi, giden havale adedi, gelen havale adedi, para yatırma toplam tutarı, şube/internet/çağrı merkezi/ivr/atm kanal kullanım adetleri alınmıştır. Bu gruptaki verilerden bazıları ay bazında değişen değerlerden oluşmaktadır. Buna, müşterinin hesap bakiye ortalamaları gibi hacim belirten veriler örnek olarak verilebilir. Bu tür verilerin anlam ifade edebilmesi için müşterinin terk etmeden önceki 6 aydaki değişimini belirleyecek şekilde bir trend değişkeni oluşturulması uygun görülmüştür. Trend değişkeni oluşturmak için bu tip alanlar için veri kümesine hem müşterinin terk etmeden önceki 1 aylık ortalama değeri, hem de terk etmeden önceki 6 aylık ortama değeri eklenecek şekilde çalışma yapılmıştır. Terk etmemiş müşteriler için ise son 1 aylık ve son 6 aylık ortalamaları trend değişkeni oluşturmak için kullanılmıştır.

Ürün sahipliği bilgileri olarak müşterinin sahip olduğu ürün sayısı ve bu ürünlerden banka için öne çıkan bazı ürünler bazında detay bilgiler alınmıştır. Kontrat bilgileri, toplam ürün/kanal sahiplikleri, mevduat ürünleri detayı, kredi kartı harcama, işlem detayları, bireysel kredi ürünü detayları, fatura ürünü detayları gibi bilgiler alınmıştır. Bu gruptaki ay bazında değişen veriler de hareket/işlem verilerinde olduğu gibi birer trend değişkenine atanarak veri kümesine dâhil edilmiştir. Bununla birlikte Ürün sahipliği verilerinde tarih belirten veriler de bulunmaktadır. Örnek olarak Kredi Kartı son hareket tarihi, Fatura Son İşlem Tarihi verilebilir.

Müşteri ile ilgili yukarıdaki bilgilere ek olarak bazı dış etkenler ile ilgili bilgiler de veri kümesine eklenmiştir. Örneğin, Müşterinin şubesi alanı dış etken olarak değerlendirilmiştir.

Terk modeli için veri kümesinin oluşturulma aşamaları Şekil 3.1 de gösterilmiştir.



Şekil 4.2 Veri Kümesinin oluşturulması

Veri Ambarında oluşturulacak yukarıdaki özelliklere sahip veriyi, operasyonel veri tabanlarından besleyecek şekilde bir ETL çalışması yapılmıştır. Veri kümesinde kullanılacak verilerin belirlenmesinden sonra ETL çalışmalarına başlanmış olup yaklaşık bir ay ETL üzerinde çalışılmıştır.

Hazırlanan veri kümesinde 133 tane alan bulunmakta olup, alanlar ve açıklamaları EK-A'da verilmiştir.

3.2.2 Verilerin Alınacağı Zaman Aralığının Belirlenmesi

Müşteri davranışlarının inceleneceği zaman aralığının belirlenmesi için mevcuttaki Elmas ve Altın değer segmentinde olup inaktife düşmüş müşterilerin bir yıllık davranış segmenti değişiklikleri incelenmiştir. 6 ay boyunca inaktif segmentte kalan ancak daha sonra geri dönen yani tekrar aktif davranış segmentlerinden birine dahil olan müşterilerin olduğu tespit edilmiştir. Bununla birlikte 6 aydan fazla inaktif durumda olan müşteriler de mevcuttur. Dolayısıyla inaktif segmente düştükten sonra müşterilerin tekrar banka ile irtibata geçebildikleri ve genellikle bu sürenin 6 ay

içerisinde bir zaman diliminde gerçekleştiği sonucuna ulaşılmıştır. Böylece müşterinin inaktife düştükten sonra en az 6 ay boyunca inaktifte kalması durumunda terk olarak tanımlanabileceğine karar verilmiştir.

Modelde kullanılacak veri kümesine, İnaktif segmente geçtikten sonra herhangi bir süre sonucunda tekrar aktif duruma dönen müşteriler dahil edilmeyeceklerdir. Müşterilerin inaktif segmente ilk düştükleri tarihten önceki 6 aylık davranışları inceleneceğinden yeni müşteri olmuş yani 6 aydan daha az süredir müşteri olanlar da veri kümesine dahil edilmeyeceklerdir.

Müşteri davranışlarının inceleneceği zaman dilimi başlangıcı olarak Aralık 2013 öncesindeki veriler dikkate alınacaktır. En fazla 3 yıllık veri dikkate alınacaktır. Yani terk eden müşterilerin terk ettikleri tarih belirlenirken en geç Ocak 2011 tarihine kadar gidilecektir.

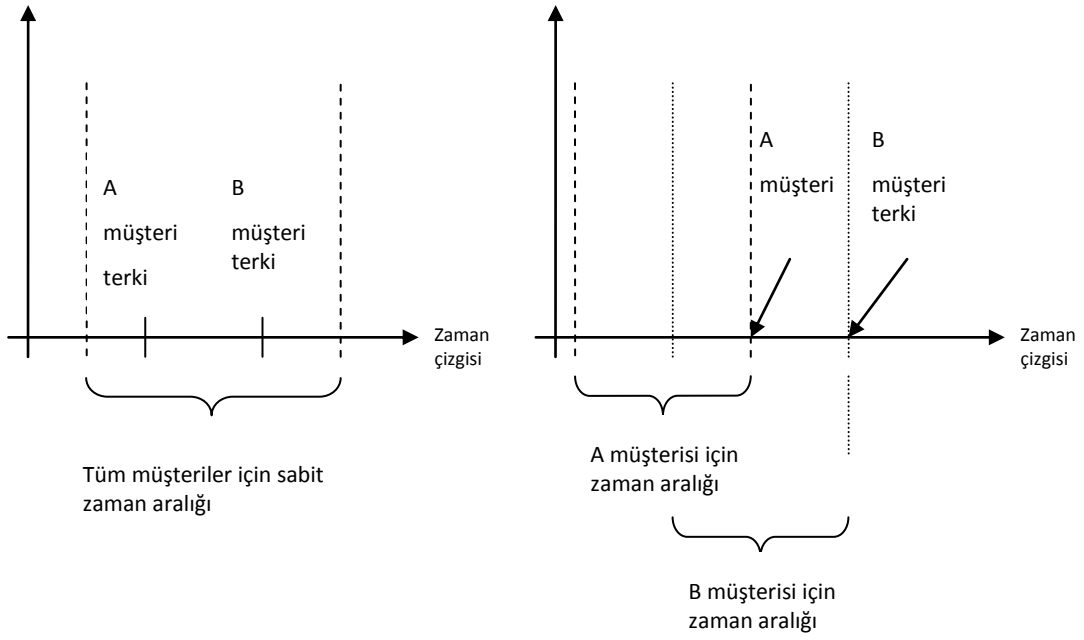
Terk eden müşterileri tahminleme modelini kurmak için, müşteri davranışlarının modellenmesi gerekmektedir. Modelin başarılı olması için de davranışlar belirli bir zaman aralığında incelenmelidir. Ancak sabit bir zaman aralığı belirlemek tam olarak ihtiyacı karşılamamaktadır. Örneğin; terk etmiş müşteriler için Ocak 2013 ile Haziran 2013 aralığındaki müşteri verileri incelendiğini varsayalım. Terk eden bazı müşterilerin Mayıs ayında, bazılarının ise Şubat ayında terk ettiklerini düşünelim. Bu durumda bazı müşteriler için terk etmeden önceki 5 aylık davranışları, bazı müşteriler için ise 1 aylık davranışları inceleniyor olacaktır. Bununla birlikte çok daha önce terk etmiş müşteriler için ise belirlenen alanların çoğunda herhangi bir değer bulunmayacaktır.

Bu durumun örneklenebilmesi için Haziran 2013 ve Ocak 2013 tarihleri arasında 6 aydan oluşan belirli bir zaman aralığı alınarak terk eden ve terk etmeyen müşterilerden oluşmuş Elmas veAltın segmentlerini içeren örnek bir veri seti alınmıştır. Terke etkisi olması beklenen alanların terk alanı ile korrelasyon analizi sonuçları incelenmiştir. Yukarıda belirtildiği gibi veri setindeki her terk eden müşteri için davranışlarının incelendiği zaman aralığı yeterli olmadığından korrelasyon olması beklenen alanların terk ile ilişkisiz ya da korrelasyon değerlerinin çok düşük çıktığı görülmüştür.

Çizelge 4.5 Sabit zaman aralığı için bazı alanların korrelasyon analiz sonucu

Pearson Correlation Coefficients, N = 80367 Prob > r under H0: Rho=0		
	TERK_F	p-değeri
CARI_HESAP	-0.00707	0.0451
TOPLAM_MEV	-0.01117	0.0015
GIDEN_EFT	-0.02496	<.0001
GELEN_EFT	-0.04258	<.0001
GIDEN_HAVALI	0.04723	<.0001
GELEN_HAVALI	-0.02999	<.0001
VIRMAN_ADET	-0.04366	<.0001
SUBE_KULLANIM	-0.05564	<.0001
INT_KULLANIM	-0.03018	<.0001
CC_KULLANIM	-0.00542	0.1244
IVR_KULLANIM	-0.00111	0.7521
ATM_KULLANIM	-0.03592	<.0001
E_COMMERCE	-0.03087	<.0001
MUSTERI_LIMIT	-0.01486	<.0001
EKSTRE_ODEME	-0.01035	0.0033
DONEM_HR	-0.01213	0.0006
TAKSITLI_ISLEM	-0.03393	<.0001
TAKSITLI_ISLEM_TUTAR	-0.01644	<.0001
E_COMMERCE_HR_TUTAR	-0.01772	<.0001
PARA_YATIRMA_ADET	-0.02948	<.0001
PARA_YATIRMA_TOP_TUTAR	-0.00374	0.2894

Bu da modelin tam olarak eğitilemeyeceği anlamına gelmektedir. Bu nedenle sabit bir zaman aralığı belirlemektense, her müşteri için değişen dinamik bir zaman aralığındaki veriler alınarak müşterilerin davranışları incelenmelidir [40].



Şekil 4.3 Sabit zaman aralığı ve dinamik zaman aralığı [40]

Bu tezdeki modelde, terk eden her müşteri için ilk inaktife düştüğü tarihten önceki 6 aylık dönemdeki verileri alınacaktır. Terk etmeyen müşterileri için ise belirlenen tarih aralığına en yakın 6 aylık dönem yani Haziran 2013-Aralık 2013 zaman aralığı seçilmiştir.

3.2.3 Veri Kümesinin Analiz Edilmesi ve Model İçin Veri Hazırlama

Veri kümesi hazırlanırken belirlenen terk tanımına uyan müşterilerin TERK_F alanı 1 olarak, diğer müşterilerin ise 0 olarak doldurulmuştur. Çizelge 3.5’de terk eden ve terk etmeyen müşterilerin frekans ve yüzdelik değerleri sunulmuştur.

Çizelge 4.6 Veri kümesindeki terk eden müşterilerin frekansı

TERK_F				
TERK_F	Frekans	Yüzde Oranı	Artımsal Frekans	Artımsal Yüzde Oranı
0	44254	94.42	44254	94.42
1	2616	5.58	46870	100

Veri kümesindeki hareket, ürün/kanal sahipliği ve kullanımı ile ilgili verilerin müşteri davranışlarındaki değişime etkisinin gözlemlenebilmesi için hem 1 aylık ortalamaları hem de 6 aylık ortalamaları veri kümesine dahil edilmiştir. Terk eden müşteriler için terk tarihinden önceki 1 aylık ve 6 aylık ortalamalar, terk etmeyen müşteriler içinse verinin alınmaya başladığı tarihten önceki 1 ayın ve önceki 6 ayın ortalaması hesaplanacak şekilde oluşturulmuştur. Bu tip verilerde her iki alanın da modele dahil edilmesinden bu alanlardan oluşturulacak bir trend değişkenini modele dahil etmek daha iyi sonuç verecektir.

- Terk eden müşteriler için terk tarihinden önceki 1 aylık verisi ile terk tarihinden önceki 6 aylık verileri kullanılarak trend değişkeni hesaplanmıştır.
- Terk etmeyen müşteriler için ise verinin alınmaya başladığı tarihten önceki 1 aylık verisi ile önceki 6 aylık verisi kullanılarak trend değişkeni hesaplanmıştır.

Örneğin Gelen_EFT_Adet_T1 ve Gelen_EFT_Adet_T2 değişkenleri için trend değişkeni şu şekilde hesaplanmıştır.

$$\text{Gelen_EFT_Adet_TR} = \left(\frac{\text{Gelen_EFT_Adet_T2} - \text{Gelen_EFT_Adet_T1}}{\text{Gelen_EFT_Adet_T2}} \right) \quad (3.1)$$

Diğer T1 ve T2 anlarındaki değişkenler için de aynı şekilde trend değişkeni hesaplanmış olup toplam 44 tane trend değişkeni oluşturulmuştur. Trend değişkenlerinin listesi EK-A'da verilmiştir.

Trend değişkeni hesaplanırken T1 ve T2 değerleri kayıp veya 0 olanlar için ise trend hesaplamasından önce ayrı bir çalışma yapılmıştır. Trend değişkeni hesaplanacak alanların tümü için kayıp ve 0 olan kayıtların ilgili alanları 0.01 olarak değiştirilmiştir. Daha sonra yukarıda belirtilen formüllere göre trend değişkenleri hesaplanmıştır.

Veri kümesinde müşterinin son aktiflik tarihi, fatura son işlem tarihi gibi tarihsel bilgiler içeren değişkenler de bulunmaktadır. Bu tür verilerin anlam ifade etmesi için de gün sayısına çevrilerek veri setine dâhil edilmesi uygun olacaktır. Gün sayıları aşağıdaki iki şekilde belirlenmiştir:

- Terk eden müşteriler için ilgili tarih verisinin, müşterinin terk tarihine kadar olan gün sayısı. Örneğin; KK_Son_Hareket_Tarihi değişkeni terk etmiş müşteriler için gün sayısı aşağıdaki formül ile hesaplanmıştır.

$$KK_Son_Aktiflik_Tarihi_DD = Terk_Tarihi - KK_Son_Aktiflik_Tarihi \quad (3.2)$$

- Terk etmeyen müşteriler için; veri setinin bitiş tarihi olan Aralık 2013'e kadar olan gün sayısı. Örneğin; KK_Son_Hareket_Tarihi değişkeni terk etmiş müşteriler için gün sayısı aşağıdaki formül ile hesaplanmıştır.

$$KK_Son_Aktiflik_Tarihi_DD = Bitis_Tarihi_Tarihi - KK_Son_Aktiflik_Tarih \quad (3.3)$$

Çizelge 4.7 Gün sayısı hesaplanmış değişkenler

Tarihsel Değişkeni	Gün Sayı Değişkeni
FATURA_ACILIS_TARIHI	FATURA_ACILIS_TARIHI_DD
FATURA_SON_ISLEM_TARIHI	FATURA_SON_ISLEM_TARIHI_DD
KK_ACILIS_TARIHI	KK_ACILIS_TARIHI_DD
KK_KONTRAT_STATU_TARIHI	KK_KONTRAT_STATU_TARIHI_DD
KK_SON_AKTIFLIK_TARIHI	KK_SON_AKTIFLIK_TARIHI_DD
KK_SON_HAREKET_TARIHI	KK_SON_HAREKET_TARIHI_DD
KK_TAKSITLER_DAHIL_SON_AKTIFLIK	KK_TAKSITLER_DAHIL_SON_AKTIFLIK_DD
KTLM_HESABI_MAX_TEMDIT_TARIHI	KTLM_HESABI_MAX_TEMDIT_TARIHI_DD
SON_AKTIFLIK_TARIHI	SON_AKTIFLIK_TARIHI_DD

Tarihsel veriler için gün sayıları hesaplandıktan sonra yine SAS EG üzerinden korrelasyon analizi sonuçları incelenmiştir. Terk _F ile SON_AKTIFLIK_TARIHI_DD, KK_ACILIS_TARIHI_DD, KK_SON_HAREKET_TARIHI_DD, KK_SON_AKTIFLIK_TARIHI_DD, KK_TAKSITLER_DAHIL_SON_AKTIF_DD, KTLM_HESABI_MAX_TEM_TAR_DD değişkenlerinin anlamlı korelasyona sahip olduğu görülmüştür.

Çizelge 4.8 Gün sayısı değişkenlerinin Terk_f alanı ile korrelasyon analizi sonuçları

Pearson Correlation Coefficients Prob > r under H0: Rho=0		
	TERK_F	
	p-value	
SON_AKTIFLIK_TARIHI_DD	-0.41275	<.0001
KK_ACILIS_TARIHI_DD	0.02851	<.0001
KK_KONTRAT_STATU_TARIHI_DD	0.01777	0.0085
KK_SON_HAREKET_TARIHI_DD	-0.16594	<.0001
KK_SON_AKTIFLIK_TARIHI_DD	-0.17916	<.0001
KK_TAKSITLER_DAHIL_SON_AKTIF_DD	-0.17916	<.0001
FATURA_SON_ISLEM_TARIHI_DD	-0.1343	<.0001
FATURA_ACILIS_TARIHI_DD	0.00697	0.2967
KTLM_HESABI_MAX_TEM_TAR_DD	-0.04484	<.0001

Bu işlemlerden sonra veri kümesi, müşterilerin detaylı hareket ve ürün/kanal kullanım verilerini içerek şekilde oluşturulmuştur. Alanların tümü SAS EG aracılığı ile analiz edilmiştir. Demografik, Nümerik ve Tarihsel olarak gruplanmış verilerin karakter analizi sonuçları EK-B’de sunulmuştur.

3.3 Model Girdilerinin Seçilmesi

Veriler analiz edilip, gerekli dönüşümler yapıp, yeni alanlar hazırlandıktan sonra model için girdi olacak değişkenlerin seçilmesi gerekmektedir. Değişkenler seçilirken veri analizinde yapılan müşteri terkinin gösteren alanla olan korrelasyon sonuçları yanında değişkenlerin birbirleri ile olan korrelasyonları ve terke etkisi de SAS EG kullanılarak incelenmiştir.

Terk_F alanı ile korrelasyona sahip olmayan değişkenler, diğer değişkenler ile yüksek korelasyonda olan değişkenler, düzgün dağılım göstermeyen değişkenler elenmiştir. Bazı değişkenler ise veri kümesine analiz ve inceleme amaçlı dahil edildiğinden bu değişkenler de modele girdi olarak alınmamıştır .

Yukarıda belirtilen sebepler nedeniyle elenen değişkenler sonucu modele girdi olarak alınan değişkenlerin listesi Çizelge 3.8’da listelenmiştir.

Çizelge 4.9 Model girdi listesi

NO	DEĞİŞKEN ADI	ROL
1	ID	ID
2	ATM_KULLANIM_ADET_TR	INPUT
3	BK_ORTALAMA_TR	INPUT
4	CALISMA_SEKLI	INPUT
5	CC_KULLANIM_ADET_TR	INPUT
6	CINSIYET	INPUT
7	COCUK_SAYISI	INPUT
8	DONEM_HARCAMA_TUTAR_TR	INPUT
9	EGITIM_DURUMU	INPUT
10	EV_SAHİPLİĞİ	INPUT
11	FATURA_SON_ISLEM_TARIHI_DD	INPUT
12	GELEN_EFT_ADET_TR	INPUT
13	GELEN_HAVALE_ADET_TR	INPUT
14	GELIRI	INPUT
15	GIDEN_EFT_ADET_TR	INPUT
16	GIDEN_HAVALE_ADET_TR	INPUT
17	INT_KULLANIM_ADET_TR	INPUT
18	IVR_KULLANIM_ADET_TR	INPUT
19	KK_EN_KOTU_ODEME_PERFORMANSI	INPUT
20	KK_IPTAL_F	INPUT
21	KK_TAKSİTLER_DAHİL_SON_AKTİF_DD	INPUT
22	KTLM_HESABI_EN_UZUN_VADE_TR	INPUT
23	KTLM_HESABI_MAX_KAR_PAYI_TR	INPUT
24	KTLM_HESABI_SON_HAREKET_TAR_DD	INPUT
25	MAAS_ODEME_ORTALAMA_TR	INPUT
26	MEDENI_DURUMU	INPUT
27	MUSTERİLİK_YAŞI	INPUT
28	MUSTERİ_LIMIT_TR	INPUT
29	PARA_YATIRMA_TOPLAM_ADET_TR	INPUT
30	SAHIP_SUBE_KOD	INPUT
31	TAKSİTLİ_ISLEM_ADET_TR	INPUT
32	TOPLAM_ETKİN_URUN_ADET_TR	INPUT
33	TOPLAM_HESAP_BAKİYE_TR	INPUT
34	TOPLAM_KANAL_SAHİPLİK_ADET_TR	INPUT
35	TOPLAM_URUN_SAHİPLİK_ADET_TR	INPUT
36	TOPLAM_VERİM	INPUT
37	UNVAN	INPUT
38	VİRMAN_ADET_TR	INPUT
39	YAKINLIK	INPUT
40	TERK_F	TARGET

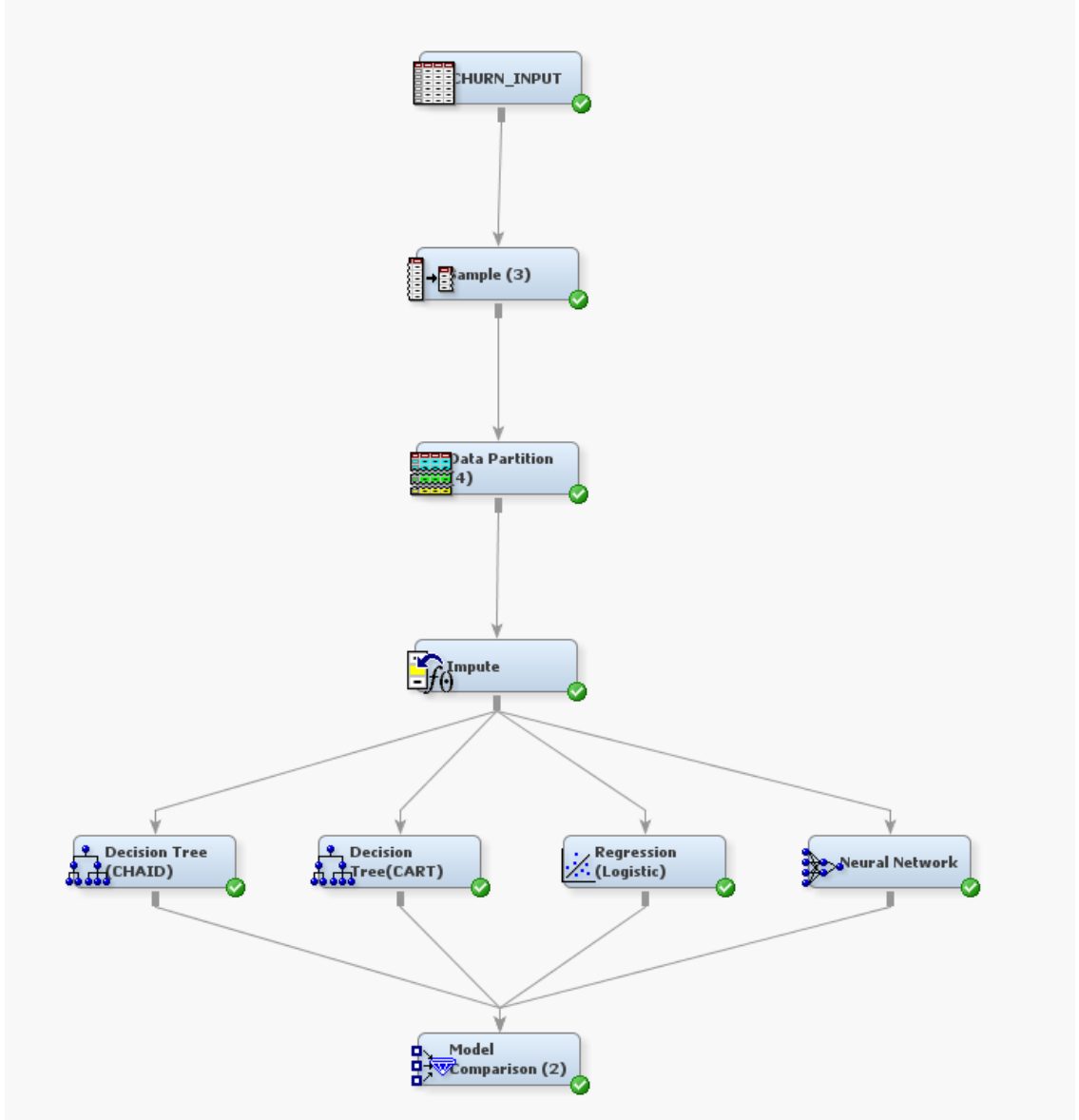
MODELİN KURULMASI**4.****4.1 Klasik Veri Madenciliği Yöntemleri İle Terk Modelleri**

Veri kümesi hazırlandıktan sonra SAS EM üzerinde veri madenciliğinin klasik modelleri olan karar ağaçları, yapay sinir ağları ve lojistik regresyon kullanılarak terk tahmin modeli geliştirilmiştir.

Model veri kümesi 46.870 müşteri verisinden oluşmuş olup, 2616 müşteri terk olarak belirlenmiştir. Etkili bir terk modeli için terk eden ve terk etmeyen müşteriler arasındaki oran yüzde 50 olacak şekilde SAS EM üzerindeki Sample nodu kullanılarak veri kümesinden bir örnek küme alınmıştır. Buna göre, örnek veri kümesinde 2616 terk ve 2616 terk etmeyen müşteri bulunmaktadır.

Data=DATA				
Variable	Numeric Value	Formatted Value	Frequency Count	Percent
TERK_F	0	0	44254	94.4186
TERK_F	1	1	2616	5.5814
Data=SAMPLE				
Variable	Numeric Value	Formatted Value	Frequency Count	Percent
TERK_F	0	0	2616	50
TERK_F	1	1	2616	50

Şekil 4.1 SAS EM Sample Nodu Sonucu



Şekil 4.2 SAS EM Üzerinde Geliştirilmiş Terk Modeli Diagramı

Daha sonra SAS EM'deki Data Partition nodu ile veri kümesinin yüzde 80'i eğitim için, yüzde 20 test için kullanılmak üzere ayrıştırılması sağlanmıştır.

Veri ayrıştırma işlemi yapıldıktan sonra verideki kayıp değerler, ilgili değişkendeki en sık görülen değer ile değiştirilmiştir.

Bu aşamalardan sonra karar ağaçları, yapay sinir ağları ve lojistik regresyon modelleri kurularak sonuçları incelenmiştir.

4.1.1 Karar Ağaçları

SAS EM üzerindeki karar ağacı nodu CART, CHAID, C4.5 ve C5.0 ağaç algoritmalarını uygulayabilecek şekilde geliştirilmiştir. CART ve CHAID en çok kullanılan algoritmalarıdır. Bölünme kuralı değiştirilerek kullanılacak metod seçilebilmektedir. SAS EM karar ağacı nodu, varsayılan olarak CART algoritmasını uygular; ağaç geliştirilirken CART'a yaklaşmak için seçilmesi gereken özellikler [41] de aşağıdaki şekilde verilmiştir.

- Bölme kuralı olarak Gini seçilmelidir.
- Maksimum dal sayısı olarak 2 seçilmelidir.
- Kayıp veriler özelliği Largest Branch olarak seçilmez.
- Alt ağaç metod özelliği Assessment olarak seçilmelidir.
- Maksimum derinlik 6 seçilmelidir.
- Yaprak boyutu 5 seçilmelidir.

Yukarıda belirtilen şekilde özellikler seçilerek CART algoritmasına yaklaşan bir KA modeli kurulmuş ve KA-CART olarak adlandırılmıştır. KA-CART 'nın test kümesindeki karmaşıklık matrisi sonuçları Çizelge 4.1'de verilmiştir.

Çizelge 4.1 KA-CART modelinin karmaşıklık matrisi

Test Kümesi (CART)					
	1	0	Toplam	Hata Oranı	Doğruluk
1	489	49	538	0.1395	0.8605
0	97	411	508		
Toplam	586	460	1046		

Model sonucu girdi değişkenlerinin önem değerleri model tarafından hesaplanır. Bu değişkenler terki en çok etkileyen kritik değişkenlerdir. Karar sayısı ise değişkenin ağaç içerisinde kaç kez yer aldığını göstermektedir. KA-CART ağacı için değişkenlerin önem değerleri Çizelge 4.2'de verilmiş olup ağacın şekilsel olarak görüntüsü EK-B'de detaylı olarak sunulmuştur.

Çizelge 4.2 KA-CART Modeli değişkenlerin önem değerleri

Değişken	Önemi	Karar Sayısı
TOPLAM_URUN_SAHİPLİK_ADET_TR	1	1
TOPLAM_HESAP_BAKIYE_TR	0.85244	2
MUSTERI_LIMIT_TR	0.84763	2
BK_ORTALAMA_TR	0.47959	1
GELEN_EFT_ADET_TR	0.33682	1
TAKSİTLİ_İSLEM_ADET_TR	0.28842	2
MUSTERİLİK_YAŞI	0.20833	1

SAS EM karar ağacı nodu seçilen özelliklere göre CART'a yaklaşılabildiği gibi, CHAID'e de yaklaşılabilecek şekilde geliştirilebilir. Aşağıdaki şekilde özellik seçimi yapılarak CHAID'e yaklaşan bir KA modeli geliştirilebilir [41].

- Bölme kuralı olarak Chi-Square seçilmelidir.
- Maksimum dal sayısı olarak veri kümesindeki kategorik değişkenlerin kategori sayılarının maksimumu seçilmelidir.
- Chi-Square Significance Level olarak 0.05 seçilmelidir.
- Alt ağaç metot özelliği Largest olarak seçilmelidir.
- Yaprak boyutu 1 seçilmelidir.

Yukarıda belirtilen şekilde özellikler seçilerek CHAID algoritmasına yaklaşan bir KA modeli kurulmuş ve KA-CHAID olarak adlandırılmıştır. KA-CHAID'in test kümesindeki karmaşıklık matrisi sonuçları Çizelge 4.3'de verilmiştir.

Çizelge 4.3 KA-CHAID Modeli karmaşıklık matrisi

Test Kümesi (KACHAID)					
	1	0	Toplam	Hata Oranı	Doğruluk
1	476	62	538	0.1290	0.8710
0	73	435	508		
Toplam	549	497	1046		

KA-CHAID ağacı için değişkenlerin önem değerleri Çizelge 4.4'de verilmiş olup şekilsel olarak görüntüsü EK-B'de detaylı olarak sunulmuştur.

Çizelge 4.4 KA-CHAID Modeli değişkenlerin önem değerleri

Değişken	Önemi	Karar Sayısı
TOPLAM_URUN_SAHİPLİK_ADET_TR	1	1
GELEN_EFT_ADET_TR	1	0.34847
TOPLAM_HESAP_BAKIYE_TR	3	0.30671
DONEM_HARCAMA_TUTAR_TR	2	0.30094
BK_ORTALAMA_TR	1	0.29995
VIRMAN_ADET_TR	1	0.19616
TOPLAM_ETKİN_URUN_ADET_TR	1	0.05181
KTLM_HESABI_MAX_KAR_PAYI_TR	2	0.03615
KK_EN_KOTU_ODEME_PERFORMANSI	1	0.02129
MUSTERI_LIMIT_TR	1	0.02090

4.1.2 Yapay Sinir Ağları

Aynı girdi değerleriyle yine SAS EM modülü üzerinde YSA nodu kullanılarak terk modeli geliştirilmiştir.

Model sonu karmaşıklık matrisi Çizelge 4.5’de verilmiştir.

Çizelge 4.5 YSA modelinin karmaşıklık matrisi

Test Kümesi (NN)					
	1	0	Toplam	Hata Oranı	Doğruluk
1	451	87	538	0.1405	0.8595
0	60	448	508		
Toplam	511	535	1046		

YSA modelinin 0,1405 hatalı sınıflandırma oranı ile terk tahmininde % 85.95 oranında doğruluk gösterdiği görülmüştür.

4.1.3 Lojistik Regresyon

KA ve YSA modellerinden sonra lojistik regresyon(Log-Reg) ile de terk tahmini modeli geliştirilmiş olup, terk tahmini için Log-Reg modelinin karmaşıklık matrisi sonuçları Çizelge 4.6’da gösterilmiştir.

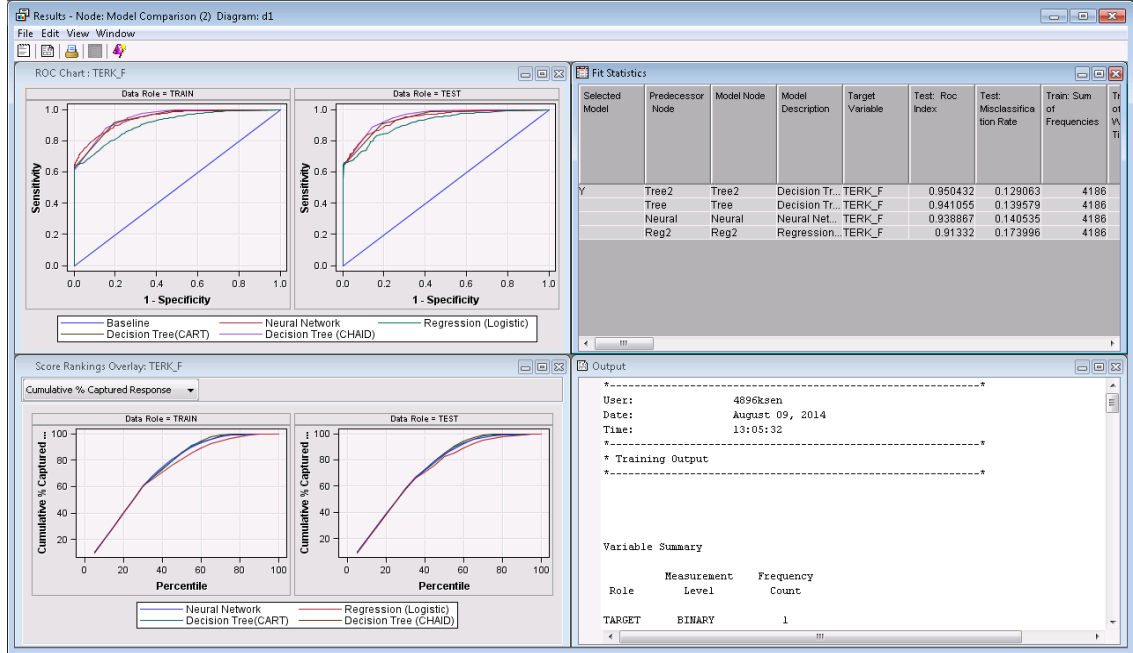
Çizelge 4.6 Lojistik regresyon modelinin karmaşıklık matrisi

Test Kümesi (Log-Reg)					
	1	0	Toplam	Hata Oranı	Doğruluk
1	430	108	538	0.1739	0.8261
0	74	434	508		
Toplam	504	542	1046		

Log-Reg modelinin 0,1890 hatalı sınıflandırma oranı ile terk tahmininde % 82.61 oranında doğruluk gösterdiği görülmüştür.

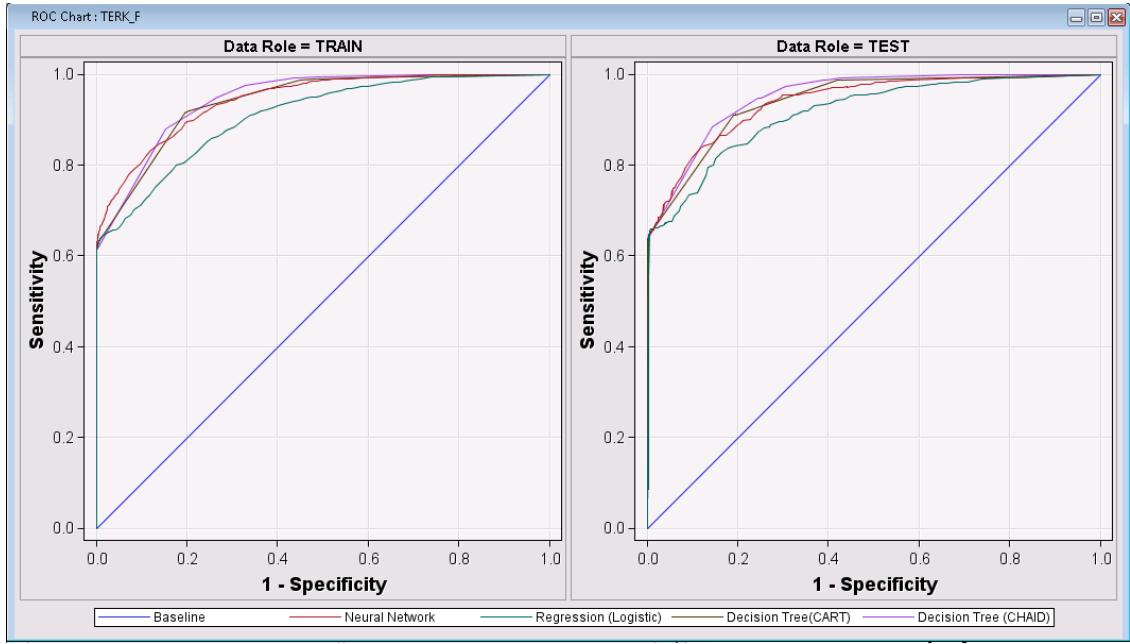
4.2 Klasik Veri Madenciliği Modellerinin Terk Modeli Karşılaştırması

SAS EM modülü üzerindeki Model Comparison nodu kullanılarak bu 4 modele ait istatistiki sonuçların karşılaştırmaları yapılmıştır. Model Comparison nodu seçilen bir istatistik sonucuna göre modelleri karşılaştırarak en iyi modelin seçilmesini sağlamaktadır.



Şekil 4.3 SAS EM Model Karşılaştırma nod Sonuçları

Model Comparison nodunda test setleri üzerinden ROC eğrisi sonuçları karşılaştırılarak sonuç yapılması şeklinde seçim yapılmıştır. Karşılaştırma sonucu 4 modelin ROC eğrisi sonucu Şekil 4.4'te verilmiştir.



Şekil 4.4 SAS EM ROC Eğrisi Karşılaştırma Sonuçları

ROC grafiğinde baseline eğrisi tahmin gücü olmayan bir modeli temsil eder. Eğri, spesiflik değerinin 0, hassaslık değerinin 1 olduğu ideal noktaya yaklaşması tahmin gücünün artması demektir. Bu nokta önemlidir, çünkü bu noktada yanlış negatif ve yanlış pozitif sayısı 0'dır. Bu nedenle AUROC değeri ne kadar 1'e yakınsa modelin tahmin gücü o kadar iyidir. Yukarıda bahsedilen 4 modelin ROC eğrisi AUROC değerleri ve test kümesi hatalı sınıflandırma sonuçları Çizelge 4.7 'de verilmiştir.

Çizelge 4.7 Modellerin AUROC ve Doğruluk Oranı Karşılaştırması

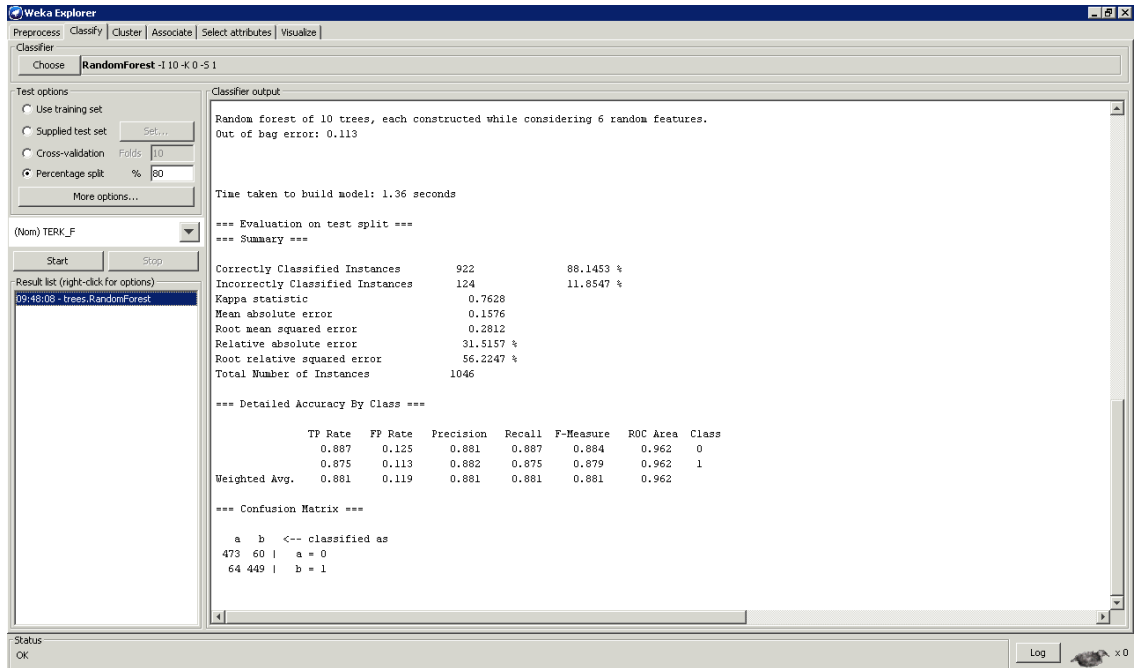
Model	AUROC	Hata Oranı	Doğruluk Oranı
KA-CHAID	0.9504	0.1290	0,8710
KA-CART	0.9401	0.1395	0,8605
YSA	0.9368	0.1405	0,8595
Log-Reg	0.9133	0.1739	0,8261

AUROC değerlerine göre karşılaştırma yapıldığında klasik veri madenciliği teknikleri arasından bu veri kümesi için en iyi sonucu model olarak KA-CHAID modeli vermiştir.

4.3 Rastgele Orman Yöntemi İle Terk Modeli

Klasik veri madenciliği yöntemleri kullanılarak terk modelleri geliştirilmiş ve modeller arasında veri kümemiz için en iyi sonucu KA CHAID'in verdiği gözlemlenmiştir. Ancak son zamanlarda özellikle bankacılık terk modellerinde, Rastgele Orman(RO) yöntemi sınıflandırmada hangi değişkenlerin önemli olduğu konusunda tahmin verebilmesi, değişken azaltmadan binlerce değişkenle çalışabilmesi, yüksek doğruluk oranı, verinin büyük çoğunluğu eksik olsa bile iyi sonuç verebilmesi, ağaç sayısı ne kadar arttırılsa da overfit olmaması gibi özelliklerinden dolayı öne çıkmaktadır. Aynı girdi değişkenleri ve aynı veri kümesi üzerinde WEKA 3.6 paket programında RO algoritması kullanılarak bir model daha geliştirilmiştir. Yine klasik KA, YSA ve Log-Reg terk modellerinde olduğu gibi veri kümesi %80 eğitim ve %20 test kümesi olacak şekilde ikiye bölünmüştür. Metot 10, 30, 50 ve 100 olacak şekilde farklı ağaç sayıları için birden fazla kere çalıştırılmış ama başarı oranında büyük bir fark gözlemlenmemiştir.

Ağaç sayısı 10 seçilerek çalıştırılan RO modelinde sonuçlar Şekil 4.5'de gösterilmiştir.



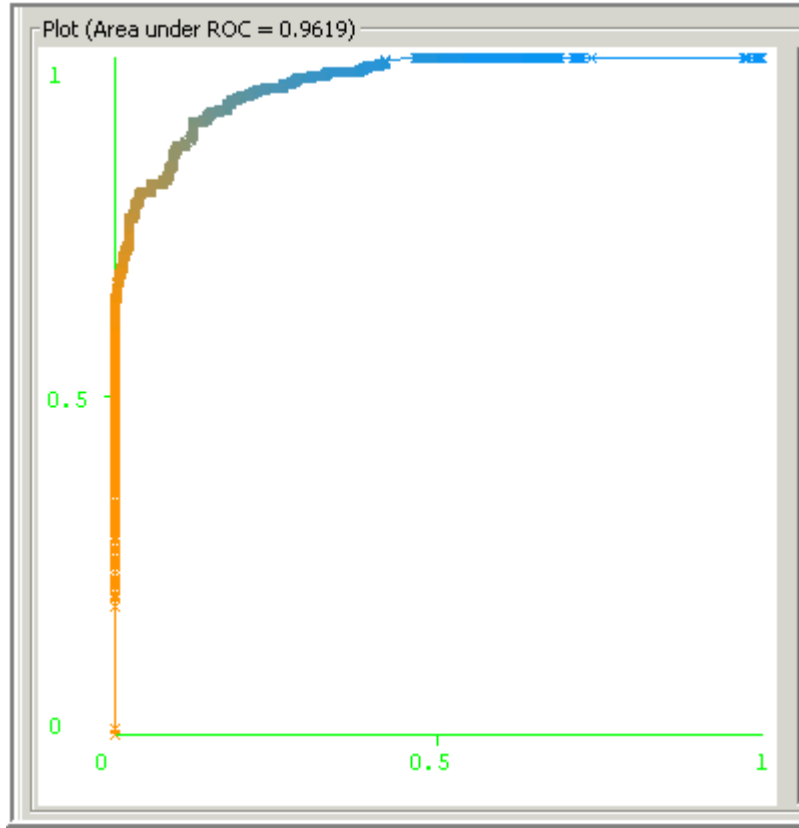
Şekil 4.5 WEKA RO Model Sonuçları

WEKA üzerinde yapılan RO tekniği ile çalıştırılmış terk modelinin karmaşıklık matrisi aşağıdaki şekildedir.

Çizelge 4.8 RO karmaşıklık matrisi

Test Kümesi (RO)					
	1	0	Toplam	Hata Oranı	Doğruluk
1	443	60	503	0.1185	0.8815
0	64	479	543		
Toplam	507	539	1046		

RO modelinin 0,1185 hatalı sınıflandırma oranı ile terk tahmininde % 88.15 oranında doğruluk gösterdiği görülmüştür. RO modelinin ROC eğrisi çizdirilmiş olup Şekil 3.6'da verilmiştir.



Şekil 4.6 RO modeli ROC eğrisi

RO için AUROC değeri 0.9619 olarak hesaplanmıştır. Klasik veri madenciliği modelleri karşılaştırıldığında RO'nun en iyi sonuç veren KA CHAID'ten daha iyi bir tahmin sonucu verdiği gözlemlenmiştir.

BÖLÜM 5

SONUÇ VE ÖNERİLER

Bu çalışmada Türkiye'deki özel bir banka için terk etmesi muhtemel müşterilerin tahmin edilmesini sağlayan bir müşteri terk tahmin modeli geliştirilmiştir. Bankacılık sektöründe müşteri ile banka arasında kontrata dayalı bir anlaşma olmadığından müşterinin tam olarak bankayı ne zaman terk ettiği bilinmemektedir. Bu nedenle tahmin modeli geliştirilebilmesi için öncelikle müşteri terk tanımına ihtiyaç vardır. Yapılan tanıma göre müşteri en az 6 ay inaktif segmentte kalmış ise terk etmiş olarak belirlenmiştir.

Bu müşterilerin önceden tahmin edilebilmesi ve müşteri kaybetmemek için gerekli önlemlerin alınabilmesi için bir model geliştirilmiştir. Modelde kullanılacak veri kümesi ürün sahipliği verileri, kanal kullanım verileri ve demografik veriler ve müşteri hareket verilerinden oluşacak şekilde 9 ayrı tablodan alınarak birleştirilmiştir. Müşterilerin bankayı terk etmeden önceki 6 aylık hareketlerinin incelenmesi amaçlanmıştır. Her müşteri için sabit bir 6 aylık dönem için veriler alındığında terk tahmini için bu verilerin yeterli olmadığı görülmüştür. Bu nedenle her müşterinin terk ettiği tarihten önceki 6 aylık dönem için hareket verileri hazırlanmış, tüm veri kümesinin zaman aralığı olarak Aralık 2013 ile Ocak 2011 yılları arasındaki veriler dikkate alınmıştır. Daha sonra SAS EG 6.1 programı kullanılarak veri kümesindeki veriler analiz edilmiş ve modele girdi olacak değişkenler seçilmiştir.

Model geliştirilirken veri madenciliği tekniklerinden yararlanılmıştır. Klasik veri madenciliği tekniklerinden olan karar ağacı, lojistik regresyon ve yapay sinir ağları

teknikleri SAS EM 6.1 programı kullanılarak geliştirilmiş olup, ek olarak WEKA 3.6 programı kullanılarak rastgele orman tekniği ile de bir terk modeli geliştirilmiştir.

Geliştirilen modeller ROC eğrileri ve doğruluk oranları üzerinden karşılaştırılmıştır. Klasik veri madenciliği teknikleri arasından CHAID algoritmasına yaklaşacak şekilde geliştirilen karar ağacı modeli en iyi sonucu vermiş olup, RO tekniği ile geliştirilen model 0.9619' luk AUROC değeri ile KACHAID modelinden de daha iyi sonuç vermiştir.

Geliştirilen veri madenciliği teknikleri sonucu modeller için veri kümesindeki en önemli değişkenlerin TOPLAM_URUN_SAHİPLİK_ADET_TR, GELEN_EFT_ADET_TR, TOPLAM_HESAP_BAKIYE_TR, DONEM_HARCAMA_TUTAR_TR, BK_ORTALAMA_TR, VIRMAN_ADET_TR, TOPLAM_ETKİN_URUN_ADET_TR, KTLM_HESABI_MAX_KAR_PAYI_TR, KK_EN_KOTU_ODEME_PERFORMANSI, MUSTERİ_LIMIT_TR olduğu görülmüştür. Bu sonuçlara istinaden terk edecek müşterilerin elde tutulması için bazı öneriler geliştirilebilir.

Müşterinin toplam ürün sahiplik adedinde azalma oluyorsa yani müşterinin terk etmeden 6 ay önceki toplam ürün sahiplik değeri terk etmeden önceki 1 ayda azalmış ise bu müşterinin terk etme olasılığı yüksektir. Bu müşteriler için çapraz satış stratejileri geliştirerek müşteriye yeni ürün satışı ile ürün sahipliğinin arttırılması yönünde kampanyalar dizayn edilebilir.

Kredi kartı ürünü bazında ise dönem harcama tutarı, kredi kartının en kötü ödeme perfomansı ve müşteri limiti önemli değişkenler olarak belirlenmiştir. Müşterinin dönem harcama tutarı, müşteri limiti azalırken terk olasılığı artmaktadır. Bu durumda limit arttırım kriterlerine uyan ancak 6 ay içinde limiti azaltılmış müşterilere limit arttırım teklifleri çıkılarak müşteri limitlerinin artırımları sağlanabilir. Dönem harcama tutarı 6 aylık dilimde azalmış müşterilere de ekstra puan teklifleri oluşturularak, kredi kartı harcamalarını arttırmaya yönelik çalışmalar yapılabilir.

Toplam hesap bakiyesi yani mevduat trendi de belirli bir seviye üzerinde azalan müşteriler için de terk olasılığı yükselmektedir. Yine mevduat ürününe yönelik olarak verilen maksimum kar payı trendindeki azalış da terk olasılığını arttırmaktadır. Bu kriterleri sağlayan müşterilerin de ilave kar payı verilerek mevduat ürünü kullanımının arttırılması sağlanabilir.

Bu alıřma, terk tahmin modeli gelecek dnemde terk edecek mřterilerin belirlenebilmesi iin kullanılacak řekilde geniřletilebilir. Veri kmesi iin yapılan ETL alıřması aylık olarak alıřacak řekilde otomatize edilerek her ay yeni dnemde terk edecek mřterilerin tahmin edilmesi saėlanabilir.

KAYNAKLAR

- [1] Blattberg, R., vd., (2008). Database Marketing Analyzing and Managing Customers, Springer Science+Business Media, LLC, New York.
- [2] Dawkins, P., ve Reichheld, F. F. (1990). "Customer retention as a competitive weapon", Directors and Boards, 14, 41-47.
- [3] Hadden, J., vd., (2006). "Churn Prediction: Does Technology Matter", International Journal of Intelligent Technology, 1(2):104–110.
- [4] Geeta, V. ve Shashi, M., (2012), "Applicability of Data Mining Techniques for Churn Prediction", The IUP Journal of Information Technology, 8 (1): 22-35.
- [5] Kitayama, M., Matsubara, R. ve Izui, Y., (2002). "Application of Data Mining to Customer Profile Analysis in the Power Electric Industry", Power Engineering Society Winter Meeting, IEEE, 1: 632 - 634.
- [6] Hwang, H., Taesoo, J. ve Euiho, S. (2004). "An LTV Model and Customer Segmentation Based on Customer Value: A Case Study on the Wireless Telecommunications Industry", Expert systems with applications, 26: 181-188
- [7] Ng, K. ve Liu, H. (2001). "Customer Retention Via Data Mining", Issues on the application of data mining, 14: 569-590.
- [8] Datta, P., Masand, B., Mani, D. R. ve Li, B. (2001). "Automated Cellular Modeling and Prediction on a Large Scale", Issues on the application of data mining, 14: 485-502.
- [9] Kim, H. ve Yoon, C. (2004). "Determinants of Subscriber Churn and Customer Loyalty in the Korean Mobile Telephony Market", Telecommunications Policy, 28: 751-765.
- [10] Mutanen T., (2006). "Customer Churn Analysis – A Case Study", Proceedings of the 2010 conference on Data Mining for Business Applications, 77-83.
- [11] Hadden J., Tiwari A., Roy R., ve Ruta D. (2006). "Churn Prediction using Complaints Data", Enformatika, 18:158.
- [12] Nath S., (2003). "Data Warehousing and Mining: Customer churn analysis in the wireless industry", Masters Abstracts International, 41(05): 1296.

- [13] Hadden J., Tiwari A., Roy R., ve Ruta D., (2007). "Computer Assisted Customer Churn Management: State-Of-The-Art and Future Trends", Computers & Operations Research, 34(10):2902-2917.
- [14] Bose I., ve Chen X., (2009). "Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn", Journal of Organizational Computing and Electronic Commerce, [http://www.informaworld.com/smpp/title~db=all~content=t775653688~tab=issueslist~branches=19 - v1919\(2\): 133 – 151, 14 Temmuz 2014](http://www.informaworld.com/smpp/title~db=all~content=t775653688~tab=issueslist~branches=19-v1919(2):133-151,14Temmuz2014)
- [15] Hung S.Y., Yen D. C., ve Wang H. Y. (2006). "Applying data mining to telecom churn management", Expert Systems with Applications, 31(3)3: 515-524.
- [16] Chu B. H, Tsai M. S., ve Ho C. S.(2007). "Toward Hybrid Data Mining Model for customer retention", Knowledge Based Systems, 20(8): 703-718.
- [17] Au, W., Chan, C. C. ve Yao, X. (2003). "A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction", IEEE transactions on evolutionary computation, 7(6): 532-545.
- [18] Rygielski, C., Wang, J. ve Yen, D. C. (2002). "Data Mining Techniques for Customer Relationship Management", Technology in Society, 24: 483-502.
- [19] Shin Yuan Hung, David C. Yen ve Hsiw Yu Wang, (2006). "Applying Data Mining To Telecom Churn Management", Expert Systems with Applications, 31(3).
- [20] Džulijana P. ve Bojana Dalbelo B., (2009). "Churn Prediction Model in Retail Banking Using Fuzzy C-Means Algorithm", Informatica, 33: 243-247.
- [21] Chiang, D., Wang, Y., ve Lee, S. and Lin, C. (2003). "Goal Orientated Sequential Pattern for Network Banking Churn Analysis", Expert systems with applications, 25: 293-302.
- [22] Madden, G. ve Savage, G. (1999). "Subscriber Churn in the Australian ISP Market", Information economics and policy, 11: 195-207.
- [23] Tao, Y. H. ve Yeh, C. R. (2003). "Simple Database Marketing Tools in Customer Analysis and Retention", International journal of information management, 23: 291-301.
- [24] Özmen M., (2006). Churn Modelling In Telecommunacation Sector, Yüksek Lisans Tezi, ITU, Fen Bilimleri Enstitüsü, İstanbul.
- [25] Tosun T., (2006). Veri Madenciliği Teknikleriyle Kredi Kartlarında Müşteri Kaybetme Analizi, Yüksek Lisans Tezi, ITU, Fen Bilimleri Enstitüsü, İstanbul.
- [26] Kışioğlu P., (2009). Telekomünikasyon Sektöründe İptal Analizi, Yüksek Lisans Tezi, ITU, Fen Bilimleri Enstitüsü, İstanbul.
- [27] Bilgen O., (2009). Churn Management of Electronic Banking Customers, Yüksek Lisans Tezi, Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [28] Koçtürk Y., (2010). Veri Madenciliğinde Bağlılık, Yüksek Lisans Tezi, ITU, Fen Bilimleri Enstitüsü, İstanbul.

- [29] Güllüoğlu B., (2011). Customer Segmentation For Churn Management By Using Ant Colony, Yüksek Lisans Tezi, Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [30] Arifoğlu E., (2011). Churn Management By Using Fuzzy C-Means, Yüksek Lisans Tezi, Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [31] Arıtürk, U., (2011). Improved Churn Prediction by More Effective Use of Customer Data: The Case of Private Banking Customers, Yüksek Lisans Tezi, Koç Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [32] SAS Institute Inc., http://www.sas.com/en_us/home.html, 14 Temmuz 2014.
- [33] WEKA, The University of Waikato, <http://www.cs.waikato.ac.nz/ml/weka/>, 14 Temmuz 2014.
- [34] Verbeke, W., Martens, D., Mues, C., ve Baesens, B. (2011). “Building comprehensible customer churn prediction models with advanced rule induction techniques”, Expert Systems with Applications, 38(3):2354–2364.
- [35] L. Breiman, (2001). “Random forests”, Machine Learning, 45: 5-32.
- [36] Verikas, A., Gelzinis A., ve Bacauskiene, M., (2010). “Mining data with random forests: A survey and results of new tests”, IEEE Pattern Recognition, 44(2): 330–349
- [37] Walker, M., Random Forests Algorithm, 2013, <http://www.datasciencecentral.com/profiles/blogs/random-forests-algorithm>, 14 Temmuz 2014.
- [38] Gladys, N., (2009), “Modeling churn using customer lifetime value”, European journal of operational research : EJOR.- Amsterdam : Elsevier, ISSN 0377-2217, ZDB-ID 2430034. - Vol. 197.2009, 1, s. 402-411.
- [39] Tiwari, A., (2007), “Computer assisted customer churn management: State-of-the-art and future trends, Computers & Operations Research.
- [40] Prasad, D., ve Madhavi, S., (2012), “Prediction of Churn Behavior of Bank Customers Using Data Mining Tools”, Business Intelligence Journal.
- [41] Matignon, R., (2007). Data Mining Using SAS Enterprise Miner, John Wiley & Sons, 638, New Jersey

VERİ KÜMESİ

A-1 Veri kümesindeki alanlar ve açıklamaları

	Değişken Adı	Açıklama
1	DEGER_SEGMENT_KOD	Müşterinin değer segment kodu
2	DAVRANIS_SEGMENT_KOD	Müşterinin davranış segment kodu
3	SAHIP_SUBE_KOD	Müşterinin bağlı olduğu şube bilgisi
4	EGITIM_DURUMU	Müşterinin eğitim durumu
5	CINSIYET	Müşterinin cinsiyeti
6	UNVAN	Müşterinin unvanı
7	MESLEK_KODU	Müşterinin meslek bilgisi
8	CALISMA_SEKLI	Müşterinin çalışma şekli
9	MEDENI_DURUMU	Müşterinin medeni durumu
10	COCUK_SAYISI	Müşterinin sahip olduğu çocuk sayısı
11	YAKINLIK	Müşterinin kuruma yakınlığı
12	MUSTERILIK_YASI	Müşterinin ay cinsinden ne kadar süredir müşteri olduğu
13	YAS	Müşterinin yaşı
14	GELIRI	Müşterinin gelir bilgisi
15	UYRUGU	Müşterinin uyruk bilgisi
16	EV_SAHİPLİĞİ	Müşterinin ev sahibi olup olmadığı bilgisi
17	ACIK_KONTRAT_ADET_T1	1 ay önceki açık kontrat adedi toplamı
18	ACIK_KONTRAT_ADET_T2	6 ay önceki açık kontrat adedi toplamı
19	KAPALI_KONTRAT_ADET_T1	1 ay önceki kapalı kontrat adedi toplamı
20	KAPALI_KONTRAT_ADET_T2	6 ay önceki kapalı kontrat adedi toplamı
21	AKTIF_KONTRAT_ADET_T1	1 ay önceki aktif kontrat adedi toplamı

22	AKTIF_KONTRAT_ADET_T2	6 ay önceki aktif kontrat adedi toplamı
23	ETKIN_KONTRAT_ADET_T1	1 ay önceki etkin kontrat adedi toplamı
24	ETKIN_KONTRAT_ADET_T2	6 ay önceki etkin kontrat adedi toplamı
25	TOPLAM_URUN_SAHİPLİK_ADET_T1	1 ay önceki toplam ürün sahiplik adedi
26	TOPLAM_URUN_SAHİPLİK_ADET_T2	6 ay önceki toplam ürün sahiplik adedi
27	TOPLAM_KANAL_SAHİPLİK_ADET_T1	1 ay önceki toplam kanal sahiplik adedi
28	TOPLAM_KANAL_SAHİPLİK_ADET_T2	6 ay önceki toplam kanal sahiplik adedi
29	TOPLAM_AKTIF_URUN_ADET_T1	1 ay önceki toplam aktif ürün adedi
30	TOPLAM_AKTIF_URUN_ADET_T2	6 ay önceki toplam aktif ürün adedi
31	TOPLAM_ETKIN_URUN_ADET_T1	1 ay önceki toplam etkin ürün adedi
32	TOPLAM_ETKIN_URUN_ADET_T2	6 ay önceki toplam etkin ürün adedi
33	KTLM_HESABI_TOPLAM_ANAPARA	Katılma hesabı toplam anaparası
34	KTLM_HESABI_KAZANILAN_KARPAYI	Katılma hesabı kazanılan kar payı
35	KTLM_HESABI_KAPAMA_TRANSFER_BANKA	Katılma hesabı kapatma transfer bankası
36	KTLM_HESABI_ILAVE_KARPAYI_F	Katılma hesabı ilave kar payı alıp almadığı bilgisi
37	KTLM_HESABI_MAX_KAR_PAYI_T1	1 ay önceki maximum kar payı değeri
38	KTLM_HESABI_MAX_KAR_PAYI_T2	6 ay önceki maximum kar payı değeri
39	KTLM_HESABI_TOPLAM_ANAPARA_T1	1 ay önceki ortalama toplam anapara değeri
40	KTLM_HESABI_TOPLAM_ANAPARA_T2	6 ay önceki ortalama toplam anapara değeri
41	KTLM_HESABI_EN_UZUN_VADE_T1	1 ay önceki katılma hesap en uzun vade tarihi
42	KTLM_HESABI_EN_UZUN_VADE_T2	6 ay önceki katılma hesap en uzun vade tarihi
43	ACIK_KTLM_HESABI_ADET_T1	1 ay önceki açık katılma hesap adedi
44	ACIK_KTLM_HESABI_ADET_T2	6 ay önceki açık katılma hesap adedi
45	KTLM_HESABI_MAX_TEMDIT_TARIHI_T1	1 ay önceki katılma hesabı max temdit tarihi
46	KTLM_HESABI_MAX_TEMDIT_TARIHI_T2	6 ay önceki katılma hesabı max temdit tarihi
47	TOPLAM_GELİR	Müşterinin bankaya toplam geliri
48	TOPLAM_GIDER	Müşterinin bankaya toplam gideri
49	TOPLAM_VERİM	Müşterinin banka için verimi
50	MUSTERI_VERİM_T1	1 ay önceki müşteri verimi
51	MUSTERI_VERİM_T2	6 ay önceki müşteri verimi
52	KK_KONTRAT_STATU_KODU	Kredi kartı kontrat statü kodu
53	KK_KONTRAT_STATU_TARIHI	Kredi kartı kontrat statü tarihi
54	KK_ACILIS_TARIHI	Kredi kartı açılış tarihi
55	KK_SON_HAREKET_TARIHI	Kredi kartı son hareket tarihi
56	KK_IPTAL_F	Kredi kartı iptal nedeni

57	KK_SON_AKTIFLIK_TARIHI	Kredi kartı son aktiflik tarihi
58	KK_TAKSITLER_DAHIL_SON_AKTIFLIK_	Kredi kartı taksitler dâhil son aktiflik tarihi
59	KK_EN_KOTU_ODEME_PERFORMANSI	Kredi kartı en kötü ödeme performansı
60	KK_SON_KULLANMA_TARIHI	Kredi kartı son kullanma tarihi
61	EKSTRE_ODEME_TUTAR_T1	1 ay önceki ekstre ödeme tutarı
62	EKSTRE_ODEME_TUTAR_T2	6 ay önceki ekstre ödeme tutarı
63	DONEM_HARCAMA_TUTAR_T1	1 ay önceki ortamala kredi kartı dönem harcama tutarı
64	DONEM_HARCAMA_TUTAR_T2	6 ay önceki ortamala kredi kartı dönem harcama tutarı
65	TAKSITLI_ISLEM_ADET_T1	1 ay önceki kredi kartı taksitli işlem adedi
66	TAKSITLI_ISLEM_ADET_T2	6 ay önceki kredi kartı taksitli işlem adedi
67	TAKSITLI_ISLEM_TUTAR_T1	1 ay önceki ortalama kredi kartı taksitli işlem tutarı
68	TAKSITLI_ISLEM_TUTAR_T2	6 ay önceki ortalama kredi kartı taksitli işlem tutarı
69	FATURA_SON_ISLEM_TARIHI	Fatura son işlem tarihi
70	FATURA_ACILIS_TARIHI	Fatura kontrat açılış tarihi
71	ACIK_KONUT_KREDISI_F_T1	1 ay önce açık konut kredisi olup olmadığı
72	ACIK_KONUT_KREDISI_F_T2	6 ay önce açık konut kredisi olup olmadığı
73	ACIK_TASIT_KREDISI_F_T1	1 ay önce açık taşıt kredisi olup olmadığı
74	ACIK_TASIT_KREDISI_F_T2	6 ay önce açık taşıt kredisi olup olmadığı
75	ACIK_TUKETICI_KREDISI_F_T1	1 ay önce açık tüketici kredisi olup olmadığı
76	ACIK_TUKETICI_KREDISI_F_T2	6 ay önce açık tüketici kredisi olup olmadığı
77	BK_TOPLAM_TAKSIT_ADET	Bireysel kredi toplam taksit adedi
78	BK_TOPLAM_RISK_T1	1 ay önceki toplam bireysel kredi riski
79	BK_TOPLAM_RISK_T2	6 ay önceki toplam bireysel kredi riski
80	ACIK_BK_ADET_T1	1 ay önceki açık bireysel kredi adedi
81	ACIK_BK_ADET_T2	6 ay önceki açık bireysel kredi adedi
82	KAPALI_BK_ADET_T1	1 ay önceki kapalı bireysel kredi adedi
83	KAPALI_BK_ADET_T2	6 ay önceki kapalı bireysel kredi adedi
84	EN_KOTU_ODEME_PERFORMANSI	Bireysel kredi en kötü ödeme performansı
85	BK_ORTALAMA_T1	1 ay önceki ortalama bireysel kredi tutarı
86	BK_ORTALAMA_T2	6 ay önceki ortalama bireysel kredi tutarı
87	CARI_HESAP_ORTALAMA_T1	1 ay önceki cari hesap ortalama bakiyesi
88	CARI_HESAP_ORTALAMA_T2	6 ay önceki cari hesap ortalama bakiyesi
89	KTLM_HESAP_ORTALAMA_T1	1 ay önceki katılma hesap ortalama bakiyesi
90	KTLM_HESAP_ORTALAMA_T2	6 ay önceki katılma hesap ortalama bakiyesi
91	TOPLAM_HESAP_BAKIYE_T1	1 ay önceki toplam hesap ortalama bakiyesi
92	TOPLAM_HESAP_BAKIYE_T2	6 ay önceki toplam hesap ortalama bakiyesi

93	MAAS_ODEME_ORTALAMA_T1	1 ay önceki ortalama maas ödeme tutarı
94	MAAS_ODEME_ORTALAMA_T2	6 ay önceki ortalama maas ödeme tutarı
95	YATIRIM_HESAP_BAKIYESI_T1	1 ay önceki ortalama hesap bakiyesi
96	YATIRIM_HESAP_BAKIYESI_T2	6 ay önceki ortalama hesap bakiyesi
97	GIDEN_EFT_ADET_T1	1 ay önceki giden EFT adedi
98	GIDEN_EFT_ADET_T2	6 ay önceki giden EFT adedi
99	GELEN_EFT_ADET_T1	1 ay önceki gelen EFT adedi
100	GELEN_EFT_ADET_T2	6 ay önceki gelen EFT adedi
101	GIDEN_HAVALE_ADET_T1	1 ay önceki giden havale adedi
102	GIDEN_HAVALE_ADET_T2	6 ay önceki giden havale adedi
103	GELEN_HAVALE_ADET_T1	1 ay önceki gelen havale adedi
104	GELEN_HAVALE_ADET_T2	6 ay önceki gelen havale adedi
105	VIRMAN_ADET_T1	1 ay önceki virman adedi
106	VIRMAN_ADET_T2	6 ay önceki virman adedi
107	PARA_YATIRMA_TOPLAM_ADET_T1	1 ay önceki toplam para yatırma adedi
108	PARA_YATIRMA_TOPLAM_ADET_T2	6 ay önceki toplam para yatırma adedi
109	PARA_YATIRMA_TOPLAM_TUTAR_T1	1 ay önceki ortalama para yatırma tutarı
110	PARA_YATIRMA_TOPLAM_TUTAR_T2	6 ay önceki ortalama para yatırma tutarı
111	E_COMMERCE_HARCAMA_TUTAR_T1	1 ay önceki ortalama e-commerce harcama tutarı
112	E_COMMERCE_HARCAMA_TUTAR_T2	6 ay önceki ortalama e-commerce harcama tutarı
113	E_COMMERCE_HARCAMA_ADET_T1	1 ay önceki ortalama e-commerce harcama adedi
114	E_COMMERCE_HARCAMA_ADET_T2	6 ay önceki ortalama e-commerce harcama adedi

4		
11		
5	SUBE_KULLANIM_ADET_T1	1 ay önceki şube kullanım adedi
11		
6	SUBE_KULLANIM_ADET_T2	6 ay önceki şube kullanım adedi
11		
7	INT_KULLANIM_ADET_T1	1 ay önceki internet kullanım adedi
11		
8	INT_KULLANIM_ADET_T2	6 ay önceki internet kullanım adedi
11		
9	CC_KULLANIM_ADET_T1	1 ay önceki call center kullanım adedi
12		
0	CC_KULLANIM_ADET_T2	6 ay önceki call center kullanım adedi
12		
1	IVR_KULLANIM_ADET_T1	1 ay önceki IVR kullanım adedi
12		
2	IVR_KULLANIM_ADET_T2	6 ay önceki IVR kullanım adedi
12		
3	ATM_KULLANIM_ADET_T1	1 ay önceki ATM kullanım adedi
12		
4	ATM_KULLANIM_ADET_T2	6 ay önceki ATM kullanım adedi
12		
5	MUSTERI_LIMIT_T1	1 ay önceki müşteri limiti
12		
6	MUSTERI_LIMIT_T2	6 ay önceki müşteri limiti
12		
7	SON_AKTIFLIK_TARIHI	Müşterinin son aktiflik tarihi
12		
8	FINANSAL_ISLEM_ADET_T1	1 ay önceki toplam finansal işlem adedi
12		
9	FINANSAL_ISLEM_ADET_T2	6 ay önceki toplam finansal işlem adedi
13		
0	FINANSAL_OLMAYAN_ISLEM_ADET_T1	1 ay önceki toplam finansal olmayan işlem adedi
13		
1	FINANSAL_OLMAYAN_ISLEM_ADET_T2	6 ay önceki toplam finansal olmayan işlem adedi
13		
2	TERK_F	Müşterinin terk edip etmediği bilgisi

13		
3	TERK_YIL_AY	Terk eden müşteriler için terk ettiği ay/yıl bilgisi

A-2 Oluşturulan Trend Değişkenleri

	Trend Değişkenleri	Değişkenler
1	ACIK_BK_ADET_TR	ACIK_BK_ADET_T1 ACIK_BK_ADET_T2
2	ACIK_KONTRAT_ADET_TR	ACIK_KONTRAT_ADET_T1 ACIK_KONTRAT_ADET_T2
3	ACIK_KTLM_HESABI_ADET_TR	ACIK_KTLM_HESABI_ADET_T1 ACIK_KTLM_HESABI_ADET_T2
4	AKTIF_KONTRAT_ADET_TR	AKTIF_KONTRAT_ADET_T1 AKTIF_KONTRAT_ADET_T2
5	ATM_KULLANIM_ADET_TR	ATM_KULLANIM_ADET_T1 ATM_KULLANIM_ADET_T2
6	BK_ORTALAMA_TR	BK_ORTALAMA_T1 BK_ORTALAMA_T2
7	BK_TOPLAM_RISK_TR	BK_TOPLAM_RISK_T1 BK_TOPLAM_RISK_T2
8	CARI_HESAP_ORTALAMA_TR	CARI_HESAP_ORTALAMA_T1 CARI_HESAP_ORTALAMA_T2
9	CC_KULLANIM_ADET_TR	CC_KULLANIM_ADET_T1 CC_KULLANIM_ADET_T2
10	DEGER_SEGMENT_KOD_TR	DEGER_SEGMENT_KOD_T1 DEGER_SEGMENT_KOD_T2
11	DONEM_HARCAMA_TUTAR_TR	DONEM_HARCAMA_TUTAR_T1 DONEM_HARCAMA_TUTAR_T2
12	EKSTRE_ODEME_TUTAR_TR	EKSTRE_ODEME_TUTAR_T1 EKSTRE_ODEME_TUTAR_T2
13	ETKIN_KONTRAT_ADET_TR	ETKIN_KONTRAT_ADET_T1 ETKIN_KONTRAT_ADET_T2
14	E_COMMERCE_HARCAMA_ADET_TR	E_COMMERCE_HARCAMA_ADET_T1 E_COMMERCE_HARCAMA_ADET_T2
15	E_COMMERCE_HARCAMA_TUTAR_TR	E_COMMERCE_HARCAMA_TUTAR_T1 E_COMMERCE_HARCAMA_TUTAR_T2
16	FINANSAL_ISLEM_ADET_TR	FINANSAL_ISLEM_ADET_T1 FINANSAL_ISLEM_ADET_T2
17	FINANSAL_OLMAYAN_ISLEM_ADET_TR	FINANSAL_OLMAYAN_ISLEM_ADET_T1 FINANSAL_OLMAYAN_ISLEM_ADET_T2
18	GELEN_EFT_ADET_TR	GELEN_EFT_ADET_T1 GELEN_EFT_ADET_T2
19	GELEN_HAVALA_ADET_TR	GELEN_HAVALA_ADET_T1 GELEN_HAVALA_ADET_T2
20	GIDEN_EFT_ADET_TR	GIDEN_EFT_ADET_T1 GIDEN_EFT_ADET_T2
21	GIDEN_HAVALA_ADET_TR	GIDEN_HAVALA_ADET_T1 GIDEN_HAVALA_ADET_T2
22	INT_KULLANIM_ADET_TR	INT_KULLANIM_ADET_T1 INT_KULLANIM_ADET_T2
23	IVR_KULLANIM_ADET_TR	IVR_KULLANIM_ADET_T1

		IVR_KULLANIM_ADET_T2
24	KAPALI_BK_ADET_TR	KAPALI_BK_ADET_T1
		KAPALI_BK_ADET_T2
25	KAPALI_KONTRAT_ADET_TR	KAPALI_KONTRAT_ADET_T1
		KAPALI_KONTRAT_ADET_T2
26	KTLM_HESABI_EN_UZUN_VADE_TR	KTLM_HESABI_EN_UZUN_VADE_T1
		KTLM_HESABI_EN_UZUN_VADE_T2
27	KTLM_HESABI_MAX_KAR_PAYI_TR	KTLM_HESABI_MAX_KAR_PAYI_T1
		KTLM_HESABI_MAX_KAR_PAYI_T2
28	KTLM_HESABI_TOPLAM_ANAPARA_TR	KTLM_HESABI_TOPLAM_ANAPARA_T1
		KTLM_HESABI_TOPLAM_ANAPARA_T2
29	KTLM_HESAP_ORTALAMA_TR	KTLM_HESAP_ORTALAMA_T1
		KTLM_HESAP_ORTALAMA_T2
30	MAAS_ODEME_ORTALAMA_TR	MAAS_ODEME_ORTALAMA_T1
		MAAS_ODEME_ORTALAMA_T2
31	MUSTERI_LIMIT_TR	MUSTERI_LIMIT_T1
		MUSTERI_LIMIT_T2
32	MUSTERI_VERIM_TR	MUSTERI_VERIM_T1
		MUSTERI_VERIM_T2
33	PARA_YATIRMA_TOPLAM_ADET_TR	PARA_YATIRMA_TOPLAM_ADET_T1
		PARA_YATIRMA_TOPLAM_ADET_T2
34	PARA_YATIRMA_TOPLAM_TUTAR_TR	PARA_YATIRMA_TOPLAM_TUTAR_T1
		PARA_YATIRMA_TOPLAM_TUTAR_T2
35	SUBE_KULLANIM_ADET_TR	SUBE_KULLANIM_ADET_T1
		SUBE_KULLANIM_ADET_T2
36	TAKSITLI_ISLEM_ADET_TR	TAKSITLI_ISLEM_ADET_T1
		TAKSITLI_ISLEM_ADET_T2
37	TAKSITLI_ISLEM_TUTAR_TR	TAKSITLI_ISLEM_TUTAR_T1
		TAKSITLI_ISLEM_TUTAR_T2
38	TOPLAM_AKTIF_URUN_ADET_TR	TOPLAM_AKTIF_URUN_ADET_T1
		TOPLAM_AKTIF_URUN_ADET_T2
39	TOPLAM_ETKIN_URUN_ADET_TR	TOPLAM_ETKIN_URUN_ADET_T1
		TOPLAM_ETKIN_URUN_ADET_T2
40	TOPLAM_HESAP_BAKIYE_TR	TOPLAM_HESAP_BAKIYE_T1
		TOPLAM_HESAP_BAKIYE_T2
41	TOPLAM_KANAL_SAHİPLİK_ADET_TR	TOPLAM_KANAL_SAHİPLİK_ADET_T1
		TOPLAM_KANAL_SAHİPLİK_ADET_T2
42	TOPLAM_URUN_SAHİPLİK_ADET_TR	TOPLAM_URUN_SAHİPLİK_ADET_T1
		TOPLAM_URUN_SAHİPLİK_ADET_T2
43	VIRMAN_ADET_TR	VIRMAN_ADET_T1
		VIRMAN_ADET_T2
44	YATIRIM_HESAP_BAKİYESİ_TR	YATIRIM_HESAP_BAKİYESİ_T1
		YATIRIM_HESAP_BAKİYESİ_T2

DEĞİŞKENLERİN KARAKTER ANALİZİ SONUÇLARI

B-1 Kategorik değişkenlerin karakter analizi sonuçları

Variable	Value	Frequency	Percent of Total
		Count	Frequency
CINSİYET	ERKEK	35315	75.347
	KADIN	11528	24.596
	Missing	27	0.058
DEGER_SEGMENT_AD_T2	Altın	28883	61.624
	Elmas	17987	38.376
EGITIM_DURUMU	***Missing***	17901	38.193
	ÜNİVERSİTE	11767	25.106
	LİSE	11343	24.201
	İLKOKUL	2080	4.438
	YÜKSEK LİSANS	1247	2.661
	ORTAOKUL	1083	2.311
	MESLEK Y.O.	965	2.059
	DOKTORA	461	0.984
	OKULA GITMİYOR	23	0.049
EV_SAHİPLİĞİ	***Missing***	24919	53.166
	KENDİSİNİN	10791	23.023
	KİRA	8149	17.386
	AİLEYE AIT	2628	5.607
	LOJMAN	383	0.817
KK_IPTAL_NEDEN_ADI	***Missing***	46505	99.221

	YENILENME DIGİNDEN	228	0.486
	Kart Ücreti Nedeniyle	133	0.284
	MÜKERRER KART	3	0.006
	GENEL MÜDÜRLÜK İPTAL (DİGER)	1	0.002
KK_KONTRAT_STATU_KODU	***Missing***	24942	53.215
	P	16413	35.018
	L	2643	5.639
	A	1833	3.911
	G	427	0.911
	K	201	0.429
	T	160	0.341
	F	140	0.299
	S	82	0.175
	U	29	0.062
KTLM_HESABI_KAPAMA_TRANSFER_BANK	***Missing***	46870	100
MEDENİ_DURUMU	EVLI	34702	74.039
	Missing	8305	17.719
	BEKAR	3863	8.242
UYRUGU	TÜRKİYE	46520	99.253
	ALMANYA	161	0.344
	SURİYE	32	0.068
	AVUSTURYA	26	0.055
	DANİMARKA	11	0.023
	RUSYA	11	0.023
	AZERBAYCAN	10	0.021
	KAZAKİSTAN	6	0.013
	SUUDİ ARABİ	6	0.013
	HOLLANDA	5	0.011
	İRAK	5	0.011
	YUNANİSTAN	5	0.011
	AVUSTRALYA	4	0.009
	BİRL. ARAP EM	4	0.009
	CİN HALK CUM	3	0.006
	FRANSA	3	0.006
	İNGİLTERE	3	0.006

	KIRGIZISTAN	3	0.006
	MISIR	3	0.006
	ÖZBEKISTAN	3	0.006
	ARNAVUTLUK	2	0.004
	BREZİLYA	2	0.004
	BULGARISTAN	2	0.004
	ISVEC	2	0.004
	JAPONYA	2	0.004
	KUVEYT	2	0.004
	LIBYA	2	0.004
	SINGAPUR	2	0.004
	SUDAN	2	0.004
	TURKMENISTAN	2	0.004
	URDUN	2	0.004
	ABD	2	0.004
	AFGANISTAN	1	0.002
	AMERIKAN SAM	1	0.002
	CEZAYIR	1	0.002
	FILISTIN	1	0.002
	HINDISTAN	1	0.002
	ISVICRE	1	0.002
	KKTC	1	0.002
	KORE C.(G.KO	1	0.002
	KORE CUM.(GÜ	1	0.002
	KORE CUM.(KU	1	0.002
	LUBNAN	1	0.002
	MAKEDONYA	1	0.002
	NORVEC	1	0.002
	PAKISTAN	1	0.002
	PORTEKİZ	1	0.002
	SİRBİSTAN	1	0.002
	SLOVAKYA	1	0.002
	TACİKİSTAN	1	0.002
	UKRAYNA	1	0.002
	UMMAN	1	0.002

	YEMEN ARAP C	1	0.002
	YUGOSLAVYA	1	0.002
YAKINLIK	9	46509	99.23
	5	142	0.303
	26	127	0.271
	7	53	0.113
	2	7	0.015
	98	7	0.015
	456	6	0.013
	546	6	0.013
	8	4	0.009
	25	2	0.004
	6	2	0.004
	256	1	0.002
	45	1	0.002
	568	1	0.002
	62	1	0.002
	72	1	0.002

B-2 Tarihsel değişkenlerin karakter analizi sonuçları

Variable	N	NMis s	Min	Mean	Median	Max
FATURA_ACILIS_TARIHI	2240 6	2446 4	21-Feb-00	17-Jun-10	28-Jan-11	10-Jul-14
FATURA_SON_ISLEM_TARIHI	2235 1	2451 9	2-Oct-03	7-Aug-13	24-Mar-14	11-Jul-14
KK_ACILIS_TARIHI	3725 6	9614	11-Mar-98	17-Oct-09	23-Mar-10	10-Jul-14
KK_KONTRAT_STATU_TARIHI	2192 8	2494 2	8-Sep-98	10-Sep-11	23-Apr-12	31-Dec-13
KK_SON_AKTIFLIK_TARIHI	2436 3	2250 7	4-Feb-05	20-Jul-13	25-Dec-13	31-Dec-13
KK_SON_HAREKET_TARIHI	2428 2	2258 8	14-Feb-99	30-Apr-13	20-Dec-13	1-Jan-14
KK_TAKSITLER_DAHIL_SON_AKTIFLIK_	2436	2250	4-Feb-05	20-Jul-13	25-Dec-13	31-Dec-

	3	7				13
KTLM_HESABI_MAX_TEMDIT_TARIHI_T 1	2400 8	2286 2	12-May- 11	8-Nov-12	19-Nov- 12	30-Nov- 12
KTLM_HESABI_MAX_TEMDIT_TARIHI_T 2	2720 8	1966 2	12-May- 11	28-Oct- 12	15-Nov- 12	30-Nov- 12
REF_YIL_AY	4687 0	0	1-Dec-13	1-Dec-13	1-Dec-13	1-Dec-13
SON_AKTIFLIK_TARIHI	4686 3	7	6-Nov-03	2-Mar-14	30-Jun-14	22-Oct-00
TERK_YIL_AY	2616	4425 4	1-Mar-12	5-Nov-12	1-Nov-12	1-Jul-13

[illegible]

```

graph TD
    Root["0: 50.4%  
1: 49.6%  
Count: 4186"] -->|< 0.82833| Node1["0: 72.3%  
1: 27.7%  
Count: 2914"]
    Root -->|>= 0.82833| Node2["0: 0.0%  
1: 100.0%  
Count: 1272"]
    Node1 -->|< -0.125| Node3["0: 96.6%  
1: 3.4%  
Count: 715"]
    Node1 -->|>= -0.125| Node4["0: 64.4%  
1: 35.6%  
Count: 2199"]
    Node3 -->|>= 200855| Node5["0: 36.8%  
1: 63.2%  
Count: 19"]
    Node3 -->|< 200855| Node6["0: 98.3%  
1: 1.7%  
Count: 696"]
    Node4 -->|>= -0.125| Node7["0: 58.8%  
1: 41.2%  
Count: 1883"]
    Node4 -->|< -0.125| Node8["0: 97.8%  
1: 2.2%  
Count: 316"]
    Node5 -->|< -2.9967| Node9["0: 10.0%  
1: 90.0%  
Count: 10"]
    Node5 -->|>= -2.9967| Node10["0: 66.7%  
1: 33.3%  
Count: 9"]
    Node7 -->|< 0.00378| Node11["0: 48.7%  
1: 51.3%  
Count: 1183"]
    Node7 -->|>= 0.00378| Node12["0: 76.0%  
1: 24.0%  
Count: 700"]
    Node11 -->|>= -0.0001| Node13["0: 44.0%  
1: 56.0%  
Count: 1078"]
    Node11 -->|< -0.0001| Node14["0: 97.1%  
1: 2.9%  
Count: 105"]
    Node12 -->|< -0.0196| Node15["0: 32.4%  
1: 67.6%  
Count: 37"]
    Node12 -->|>= -0.0196| Node16["0: 78.4%  
1: 21.6%  
Count: 663"]
    Node13 -->|>= 30.5| Node17["0: 40.6%  
1: 59.4%  
Count: 1012"]
    Node13 -->|< 30.5| Node18["0: 95.5%  
1: 4.5%  
Count: 66"]
    Node15 -->|>= -1.125| Node19["0: 5.6%  
1: 94.4%  
Count: 18"]
    Node15 -->|< -1.125| Node20["0: 57.9%  
1: 42.1%  
Count: 19"]

```

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı :KÜBRA ŞEN
Doğum Tarihi ve Yeri :13.04.1984 Sinop
Yabancı Dili :İngilizce
E-posta :kubrasen84@gmail.com

ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Lisans	Matematik	Yıldız Teknik Üniversitesi	2006
Lise	Fen-Matematik	Kadıköy Anadolu İ.H.L.	2002