



T.C.
NECMETTİN ERBAKAN
ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**BANKACILIK SEKTÖRÜNDE MOBİL
PAZARLAMA KAMPANYALARINA
YÖNELİK HEDEF KİTLE ANALİZİ**

Recep DUR

YÜKSEK LİSANS TEZİ

Endüstri Mühendisliği Anabilim Dalı

Haziran-2019
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Recep DUR tarafından hazırlanan “BANKACILIK SEKTÖRÜNDE MOBİL PAZARLAMA KAMPANYALARINA YÖNELİK HEDEF KİTLE ANALİZİ” adlı tez çalışması 17/06/2019 tarihinde aşağıdaki jüri tarafından oy birliği ile Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Başkan

Doç. Dr. Halife KODAZ

Danışman

Prof. Dr. Sabri KOÇER

Üye

Dr. Öğr. Üyesi Onur İNAN

Yukarıdaki sonucu onaylarım.

Prof. Dr. S. Savaş DURDURAN
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.



Recep Dur

Tarih: 17/06/2019

ÖZET

YÜKSEK LİSANS TEZİ

BANKACILIK SEKTÖRÜNDE MOBİL PAZARLAMA KAMPANYALARINA YÖNELİK HEDEF KİTLE ANALİZİ

Recep DUR

**Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü
Endüstri Mühendisliği Anabilim Dalı**

Danışman: Prof. Dr. Sabri KOÇER

2019, 52 Sayfa

Jüri

Prof. Dr. Sabri KOÇER

Doç. Dr. Halife KODAZ

Dr. Öğr. Üyesi Onur İNAN

Bu çalışmada, bankacılık sektöründe mobil pazarlama kampanyalarına yönelik hedef kitle analizi gerçekleştirilmiştir. Mobil pazarlama sıklıkla kullanılan bir pazarlama yöntemi olup alakasız ve sık gönderim gibi etkenlerden dolayı müşteriler mobil pazarlama bildirimlerini almaktan vazgeçebilirler. Bu sebeple gönderim yapılmak istenen hedef kitlenin seçimi önem arz etmektedir. Analizde kullanılan veri seti 100.960 müşteriye ait 26 nitelik ve müşterilerin pazarlama SMS'lerini engelleyip engellemediğine dair sınıf bilgisini içermektedir. Bilgi kazanımı ve Relief özellik seçimi yöntemleri kullanılarak alt veri setleri oluşturulmuştur. Elde edilen veri setleri lojistik regresyon, yapay sinir ağları ve destek vektör makineleri ile sınıflandırılarak algoritmaların başarımları karşılaştırılmıştır. Yapay sinir ağları diğer sınıflandırma yöntemlerine göre daha iyi sonuç vermiştir. Bu tez çalışmasının bankacılık sektöründe mobil pazarlama kampanyaları için hedef kitle seçiminde rehberlik etmesi amaçlanmaktadır.

Anahtar Kelimeler: Destek Vektör Makineleri, Hedef Kitle Analizi, Lojistik Regresyon, Mobil Pazarlama, Özellik Seçimi, Sınıflandırma, Veri Madenciliği, Yapay Sinir Ağları

ABSTRACT

MS THESIS

**TARGET GROUP ANALYSIS FOR MOBILE MARKETING CAMPAIGNS IN
BANKING INDUSTRY**

Recep DUR

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
NECMETTİN ERBAKAN UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE
IN INDUSTRY ENGINEERING**

Advisor: Prof. Dr. Sabri KOÇER

2019, 52 Pages

Jury

Prof. Dr. Sabri KOÇER

Assoc. Prof. Dr. Halife KODAZ

Asst. Prof. Dr. Onur İNAN

In this study, customer loss analysis conducted for mobile marketing campaigns in the banking sector. Mobile marketing is a frequently used marketing method and customers can opt out of receiving mobile marketing notifications due to irrelevant and frequent delivery. For this reason, it is important to choose the target group that campaign information will be sent. The data set used in the analysis includes 26 features of 100,960 customers and class information about whether customers are blocking marketing SMSs or not. Sub-data sets are created by using information acquisition and ReliefF feature selection methods. The data sets are classified by logistic regression, artificial neural networks and support vector machines and the performance of the algorithms compared. Artificial neural networks gives better results than other classification methods. This thesis is intended to guide the selection of target audience for mobile marketing campaigns in the banking sector.

Keywords: Artificial Neural Networks, Classification, Data Mining, Feature Selection, Logistic Regression, Mobile Marketing, Support Vector Machines, Target Group Analysis

ÖNSÖZ

Tez çalışmam boyunca bilgisi ve yardımları ile beni yönlendiren değerli danışmanım Prof. Dr. Sabri KOÇER'e, beni daima destekleyen sevgili eşim Yıldız DUR'a teşekkür ederim.

Recep DUR
KONYA-2019



İÇİNDEKİLER

ÖZET	iv
ABSTRACT.....	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR.....	ix
1. GİRİŞ.....	1
2. KAYNAK ARAŞTIRMASI	3
3. MATERYAL VE YÖNTEM.....	5
3.1. Mobil Pazarlama.....	5
3.1.1. Bir mobil pazarlama aracı: SMS.....	6
3.2. Veri Madenciliği ve Veri Ön İşleme Teknikleri	8
3.2.1. Verinin temizlenmesi	10
3.2.2. Verinin bütünleştirilmesi.....	10
3.2.3. Verinin indirgenmesi.....	10
3.2.4. Verinin dönüştürülmesi.....	10
3.2.5. Veri madenciliği algoritmalarının uygulanması	11
3.2.6. Sonuçlar ve değerlendirmeler	11
3.3. Özellik Seçimi Algoritmaları	11
3.3.1. Bilgi kazanımı algoritması	12
3.3.2. ReliefF algoritması.....	13
3.4. Sınıflandırma Algoritmaları	14
3.4.1. Lojistik regresyon	14
3.4.2. Yapay sinir ağları.....	15
3.4.3. Destek vektör makineleri.....	16
3.5. K-kat Çapraz Doğrulama	18
3.6. Karmaşıklık Matrisi	19
3.7. Kullanılan Yazılım Geliştirme Ortamları	21
3.7.1. WEKA.....	21
3.7.2. Azure machine learning studio	22
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA	23
4.1. Kullanılan Veri Seti	23
4.2. Özellik Seçimi	30
4.2.1. Bilgi kazanımı özellik seçimi sonuçları	30
4.2.2. ReliefF özellik seçimi sonuçları	31
4.3. Sınıflandırma	33
4.3.1. Bilgi kazanımı sınıflandırma sonuçları	33

4.3.2. Relief sınıflandırma sonuçları	37
4.3.3. Sınıflandırma sonuçlarının incelenmesi	41
4.3.4. Yapay sinir ağları üzerine uygulama	45
5. SONUÇLAR VE ÖNERİLER	47
KAYNAKLAR	48
ÖZGEÇMİŞ.....	52



SİMGELER VE KISALTMALAR

Kısaltmalar

Arff	: Attribute Relationship File Format
Azure ML	: Azure Machine Learning Studio
BI	: Business Intelligence
BK	: Bilgi Kazanımı
DVM	: Destek Vektör Makineleri
DN	: Doğru Negatif
DP	: Doğru Pozitif
GSS	: Gönderilen Sms Sayısı
KA	: Karar Ağaçları
KNN	: K-en Yakın Komşu
LR	: Lojistik Regresyon
LRA	: Lojistik Regresyon Analizi
Maks	: Maksimum
Min	: Minimum
NB	: Naif Bayes
ÖS	: Özellik Seçimi
ROA	: Rasgele Orman Algoritması
ROC	: Reciever Operating Characteristics
SMS	: Short Message Service
YN	: Yanlış Negatif
YP	: Yanlış Pozitif
YSA	: Yapay Sinir Ağları
WEKA	: Waikato Environment for Knowledge Analysis

1. GİRİŞ

Günümüzde teknolojiye yaşanan hızlı gelişimle birlikte işletmeler pazarlama faaliyetlerinde yeni arayışlar içerisine girmektedir. Özellikle mobil ve internet teknolojilerinde yaşanan gelişmeler, işletmeleri ürün tanıtımı ve satışını arttırmaya yönelik yeni çabalar içerisine sokmuştur. Bu çerçevede işletmeler tarafından kullanılan reklam gönderimlerinde mobil pazarlama sıklıkla kullanılmaktadır.

Mobil pazarlama, mobil telefonlar üzerinden hedeflenen müşterilere iletişimin sağlanmasını amaçlayan doğrudan pazarlama anlayışı içinde değerlendirilen ve internet üzerinden pazarlama yöntemlerinden sonra ortaya çıkan yeni bir pazarlama yöntemidir (Barutçu, 2008). SMS (Short Message Service/Kısa Mesaj Hizmeti) ile reklam gönderimi en sık kullanılan mobil pazarlama yöntemlerinden birisidir. SMS hizmeti neredeyse tüm mobil telefonlar tarafından desteklendiği için mobil pazarlama alanında önemli bir yer teşkil etmektedir. Mobil pazarlama amacıyla gönderilen SMS'lerin müşterilere çok sık gönderilmesi, alakasız içerikte olması gibi sebepler müşterinin memnuniyetsizliğini artırmaktadır. Bunun neticesinde müşteri gönderilen pazarlama SMS'lerini almak istemeyebilir. Müşteri, bu durumda işletmenin sağlamış olduğu reklam gönderimlerinden çıkma kanallarını kullanarak, pazarlama SMS'lerinin kendisine gönderilmemesi hakkında faydalanabilir. İşletmelerin pazarlama amacıyla yapacağı ticari iletilerde uymakla yükümlü olduğu yönetmelik 01/05/2015 tarihinde 6563 Sayılı Elektronik Ticaretin Düzenlenmesi Hakkındaki Kanun ile yürürlüğe girmiştir. Bu mevzuat çerçevesinde işletmeler müşterilere gönderdiği pazarlama SMS'lerine ret bildirimini elektronik iletişim araçlarıyla kolay ve ücretsiz olarak iletilmesini sağlamakla ve gönderdiği iletide buna ilişkin gerekli bilgileri sunmakla yükümlüdür (Elektronik Ticaretin Düzenlenmesi, 2014).

Müşterilerin reklam bildirimini almak istememesi işletmeler için müşteri kaybını ifade etmektedir. Mobil pazarlama anlamında müşteri kayıp analizi, müşterinin bilgileri ve pazarlama amacıyla gönderilen içerikleri inceleyerek terk etme ihtimali yüksek olan müşterileri tahmin etme olarak adlandırılmaktadır. Bankacılık, telekomünikasyon, sigortacılık gibi müşteri sürekliliği bulunan sektörlerde müşteri kaybı kritik bir öneme sahiptir çünkü çoğu zaman mevcut müşterilerin tutulması, yeni müşteri kazanmaya göre daha düşük maliyetli operasyonlar gerektirir (Reichheld, Sasser, 1990).

Bu tez kapsamında, özel bir bankanın müşterilerine göndermiş olduğu mobil pazarlama kampanyalarına yönelik hedef kitle analizi gerçekleştirilmiştir. Analizde

kullanılan veri seti, müşterilere ait bazı nitelikler içermektedir. Veri seti içerisinde müşterilere ait kimlik numarası, telefon numarası gibi kişisel veri bulunmamaktadır. Analizde kullanılan veri seti özellik seçimi yöntemlerinden geçirilerek niteliklerin önem dereceleri belirlenmiştir ve sınıflandırma işlemi yapılmıştır. Sınıflandırma işleminde kullanılan yöntemlerin başarımları karşılaştırılmıştır. Yapılan çalışmanın bankacılık sektöründe mobil pazarlama kampanyaları için hedef kitle seçiminde rehberlik etmesi amaçlanmaktadır.



2. KAYNAK ARAŞTIRMASI

(Asare-Frempong ve Jayabalan, 2017), müşterilerin kampanyaya katılım oranlarını tahminlemek üzerine bir çalışma yapmışlardır. Bu çalışmada YSA, Karar Ağaçları (KA), LR ve Rasgele Orman (ROA) sınıflandırma yöntemleri kullanılmıştır. ROC eğrisi ile performansı kontrol edilmiştir. En verimli sonucu ROA sınıflandırıcısının verdiği gözlemlenmiştir.

(Karaağaç, 2015), müşteri kaybı analizi üzerine yapmış olduğu çalışmasında KA ve LR yöntemlerini kullanmıştır. Özel bir bankaya ait müşteri verileri üzerinde veri seçimi, verilerin birleştirilmesi, temizlenmesi ve dönüştürülmesi gibi veri ön işleme adımları uygulanmıştır. Yapılan testlere göre yeni geliştirilen bu yöntem ile elde edilen doğruluk oranı %89 olarak görülmektedir.

(Kaynar ve ark., 2017), yapmış oldukları çalışmada DVM, Naif Bayes (NB) ve Çok Katmanlı Yapay Sinir Ağları kullanarak 3 model elde etmişlerdir. Bir firmanın müşteri bilgilerinden oluşan veri setinden %75 oranında eğitim, %25 oranında test için rastgele veri kümeleri oluşturulmuştur. Belirtilen 3 yöntem içinde tahmin başarısı en yüksek olan modelin %92,35 oran ile YSA olduğu gözlemlenmiştir. Tahmin başarısı en düşük olan model ise %77,89 oran ile DVM'dir. Veri setindeki örnek sayısının azlığı ve eksik öznitelikler sebebiyle bu modelin başarısız olduğu öngörülmüştür.

(Apampa, 2016), banka pazarlama verisiyle kampanyaya katılım oranlarının tahminlemesi üzerine bir çalışma yapmıştır. ROA, LR, NB ve KA algoritmaları kullanılarak karşılaştırma yapılmıştır. Elde edilen sonuçlar incelendiğinde ROA'nın performans olarak daha iyi olduğu gözlemlenmiştir. Performans ölçüm yöntemlerinden sınıflandırma doğruluğu, AUC, F1 skor, geri çağırma ve hassaslık değerleri kullanılmıştır.

(Coussement ve ark., 2016), telekomünikasyon sektöründeki kayıp müşteri analizinden önce yapılan veri ön işleme adımlarının sonuçlar üzerindeki etkisini incelemişlerdir. Çalışmada veri setinin %50'si eğitim, %30'u test, %20'si seçim için kullanılmıştır. Veri ön işleme teknikleri kullanıldığında tahmin başarısının %34 oranında arttığı gözlemlenmiştir. LR, YSA ve DVM modellerinin veri ön işleme adımlarıyla etkisi incelenmiştir. Sonuç olarak LR algoritmasının YSA ve DVM metoduna göre daha hızlı olduğu gözlemlenmiştir.

(Kıraç, 2012), SMS reklamlarına yönelik tüketici tutumlarını etkileyen faktörleri inceleyen bir çalışma yapmıştır. İstanbul ilinde anket yapılarak katılımcıların SMS

reklam uygulamalarına yönelik görüşleri alınmıştır. Çalışma neticesinde, reklam SMS'lerinin faydalı ve işlevsel oluşu mobil reklamlara karşı olumlu bir tutum oluşturduğu gözlenmiştir.

(Watson ve ark., 2013), mobil pazarlamaya karşı tüketici tutumlarını inceleyen bir çalışma yapmıştır. Müşterilerin mobil pazarlama faaliyetlerine karşı tutumlarının önceden negatif olmasına rağmen, akıllı telefonların zamanla yaygınlaşması ve hayatı kolaylaştırması nedeniyle müşteri tutumlarının olumlu yönde arttığı belirlenmiştir.

(Hacıhasanoğlu, 2015), Yozgat ilinde yaptığı bir araştırmada, müşterilerin bankaların mobil pazarlama ve mobil reklam çalışmalarına olan bakış açılarını araştırmıştır. Bu araştırma için anket uygulaması gerçekleştirerek 400 müşteri üzerinde uygulamıştır. SPSS programı kullanılarak anket verileri analiz edilmiştir. Yapılan analizler neticesinde müşterilerin bankalardan gelen mobil reklamlara bakış açılarının olumlu olmadığı sonucuna ulaşılmıştır.

(Kara ve ark., 2011), yaptıkları çalışmada BIST 100 endeksinin günlük kapanış fiyatının hareket yönünü tahmin etmişlerdir. Tahmin sürecinde yapay öğrenme yöntemlerinden DVM ve YSA kullanılmıştır. İki yöntem de BIST 100 endeksinin günlük kapanış fiyatının hareket yönünü tahminde başarı sağlamıştır. Ancak YSA yöntemi DVM yönteminden az da olsa daha iyi sonuç vermiştir.

(Çakır, 2008), bankacılık müşteri veri tabanından elde ettiği veriler üzerinde birden çok sınıflandırma tekniği uygulayarak, bu tekniklerin farklılıklarını, üstünlüklerini ve zayıflıklarını analiz etmiştir. Çalışma, bankacılık müşteri veri tabanından elde edilen 188 değişken ile 17.595 kayıt içeren bir veri kümesi üzerinde Clementine 8.0 yazılımı kullanılarak gerçekleştirilmiştir. Veri madenciliğinin sınıflandırma tekniklerinden Lojistik Regresyon Analizi (LRA), YSA ve C5.0 Algoritması kullanılmıştır. Uygulama sonucunda, C5.0 algoritmasının diğerlerine göre daha hızlı olduğu gözlemlenmiştir.

3. MATERYAL VE YÖNTEM

Bu bölümde mobil pazarlama, veri madenciliği, veri ön işleme teknikleri, özellik seçimi ve sınıflandırma algoritmaları hakkında bilgi verilmeye çalışılmıştır.

3.1. Mobil Pazarlama

Son yıllarda yaşanan teknolojiadaki gelişmeler, pazarlama faaliyetlerinin daha kolay, hızlı ve verimli sağlanmasına olanak sağlamıştır. Teknolojide yaşanan bu gelişmeler mobil pazarlama kavramını ortaya çıkmasına yardımcı olmuştur.

Mobil pazarlama, cep telefonları, akıllı telefonlar vb. cihazlar yoluyla müşterilere veya potansiyel müşterilere ulaşılarak kurum, ürün, hizmet, fikir, eylem ve kişilere ilişkin pazarlama faaliyetlerinin yapılmasına olanak sağlayan bir pazarlama yöntemidir. Amerikan Mobil Pazarlama Birliği (MMA-Mobile Marketing Association) ise mobil pazarlamayı, alıcılar, tüketiciler, ortaklar ve geniş boyutuyla toplum için değer ifade eden ürünlerin geliştirilmesi, dağıtımı, mübadelesi ve ilgili gruplarla iletişim kurulmasını sağlayan bir dizi işlemler ve kuruluşlardan oluşan bir faaliyet olarak tanımlamıştır (Usta, 2009). Yapılan başka bir tanıma göre ise, mobil pazarlama, mobil telefonlar yoluyla hedef müşteriler ile pazarlama iletişiminin kurulması ve tutundurma içerikli mesajların gönderilmesi olarak ifade edilmiştir.

Yapılan tanımlar incelendiğinde, mobil pazarlamanın geniş kitlelere en hesaplı ve etkin şekilde ulaşılabilen, hedef kitle ile etkileşim kurabilen yeni bir pazarlama aracı olduğu söylenebilir. Mobil pazarlamanın işletmelere sağlamış olduğu bazı kolaylıklar aşağıdaki gibi sıralanabilir (Yüce ve ark., 2012):

- Kişiselleştirilebilme: Mobil telefonların her zaman ulaşılabilir olması ve işletmelerin kullanıcının izni doğrultusunda, kullanıcıdan elde etmiş olduğu kişisel verilerle birlikte kullanıcıya özel reklamların oluşması sağlanabilmektedir.
- Düşük Maliyet: Diğer pazarlama yöntemleriyle, örneğin kitlesel medya araçlarıyla yapılan pazarlama yöntemiyle, kıyaslandığında mobil pazarlama yoluyla yapılan iletişimin maliyeti oldukça düşüktür.
- Ölçeklenebilir Olması: Mobil pazarlama sayesinde işletmeler pazarlama amacı ile gönderilen kampanyaların ne derece etkin olduğu ve ürünlerinin kullanım alışkanlıkları ile ilgili net bilgiler edinebilmektedir. Kampanyaya katılım zamanları, kampanyaya katılım oranları, kampanyaya katılan kullanıcının

demografik istatistikleri, ürün tercihleri gibi pek çok bilgi raporlanabilmekte ve yorumlanabilmektedir.

- Hızlı ve Kolay Olması: Mobil iletişim teknolojileri sayesinde mesaj, tüketiciye kısa sürede ulaştırılabilmektedir. Tüketici kolay ve hızlı bir şekilde kampanyaya katılım durumunu geribildirim yapabilmektedir.

3.1.1. Bir mobil pazarlama aracı: SMS

SMS kavramı Short Message Service kelimelerinin baş harflerinden oluşan bir kısaltmadır. Kısa Mesaj Hizmeti anlamına gelmektedir. Cep telefonları arasında düz metinden oluşan kısa mesajların gönderilip alınabilmesi özelliğine denilmektedir. Bir SMS metni 160 karakterden oluşur. Metin içerisinde harfler, rakamlar ve alfa numerik karakterler içerebilir (Barutçu ve Öztürk Göl, 2009).

SMS reklam pazarı, hızla büyümekte olan bir mobil pazarlama yöntemidir. Pazarın bu kadar hızlı gelişme göstermesinin en önemli nedenlerinden biri sahip olduğu mesaj alma ve gönderme özelliğidir. Bu özellik reklam veren firmalar için çok önemlidir. Bu sayede gönderilmiş olan reklam hakkında geri dönüşler alabilmektedir. Bu özelliğinin yanı sıra kolay, hızlı ve güvenilir olması, mobil pazarlama açısından olumlu bir amaç sağlamaktadır (Michael ve Salter, 2006).

Türkiye gibi cep telefonunun çok yaygın olarak kullanıldığı ülkelerde, mobil pazarlama ve SMS, çeşitli sektörlerde faaliyet gösteren işletmelerde yoğun şekilde kullanılmaktadır. Örneğin bankacılık sektörü mobil pazarlamayı yoğun olarak kullanan sektörler arasında yer almaktadır. Bankalar gönderilen reklam ve etkileşimli SMS'leri sayesinde müşterilerle daha verimli ve hızlı etkileşim sağlayabilmektedir.

Mobil pazarlama reklamları tüketicilerden elde edilen bilgiler çerçevesinde kişiselleştirilebilir. Örneğin bankacılık sektörü için, müşterisinin yaşına, cinsiyetine, yaşadığı şehre vb. bilgilere göre müşterinin ihtiyacını karşılayan ve ilgisini çekebilecek ve reklamlar yapılabilir. Mobil pazarlama yöntemi SMS'i üstün kılan özellikler şöyle sıralanabilir (Taşçı, 2010).

- Kolay Erişim: SMS alabilmek için akıllı telefon olması gerekmemektedir. Tüm cep telefonları SMS alabilmektedir ve büyük bir kullanıcı kitlesi bulunmaktadır.
- Maliyet: SMS reklamlarının temel maliyeti diğer pazarlama yöntemlerine göre oldukça düşüktür. Gönderilen SMS başına GSM operatörüne bedel ödenmektedir.

- Basitlik: İşletmeler açısından basit bir geliştirme ve entegrasyon ile reklam gönderimi yapılabilir. Bunun için online toplu SMS gönderim siteleri de bulunmaktadır.
- Uygunluk: SMS farklı GSM operatörlerinde ve farklı ülkelerde genelde aynı standartlarda çalışır.

SMS'in kullanım alanları içerisinde en çok kullanılan yerler; toplu SMS ve etkileşimli SMS'tir (Aksu, 2007).

- Toplu SMS: Geliştirilen uygulamalar sayesinde ve ya web sayfalarından binlerce kişiye SMS ile ulaşmak mümkündür. Mesaj gönderilen numaralar ile ilgili iletim raporları birkaç dakika içerisinde yine aynı yöntemle bilgisayar ortamına taşınmaktadır. Toplu SMS çeşitli sektörlerde, bankacılık, alışveriş, otomotiv, market, dernek, sendikalar, yoğun olarak kullanılmaktadır.
- Etkileşimli (İnteraktif) SMS: Toplu duyuru amacıyla hizmet veren SMS sistemleri aynı zamanda bir sekreter ya da satış elemanı olarak da çalışmaktadır. Tahsis edilen bir sanal etkileşimli GSM numarası vasıtası ile kullanıcı ile etkileşimde bulunulabilir. Örnek etkileşimli SMS kullanım alanları;
 - Bir banka, müşterilerine kendi hesapları ile ilgili bilgi verebilmektedir. (BAKIYE TL yaz hesap bakiyeni öğren).
 - Bir tahlil laboratuvarı, interaktif SMS sayesinde müşterilerinin tahlil sonuçlarını SMS yoluyla öğrenebilmelerini sağlamaktadır. (TAHLIL SONUC yaz gönder)
 - Bir işletme SMS ile sipariş alabilir. (SIPARIS ONAY ya da SIPARIS IPTAL, yaz gönder)

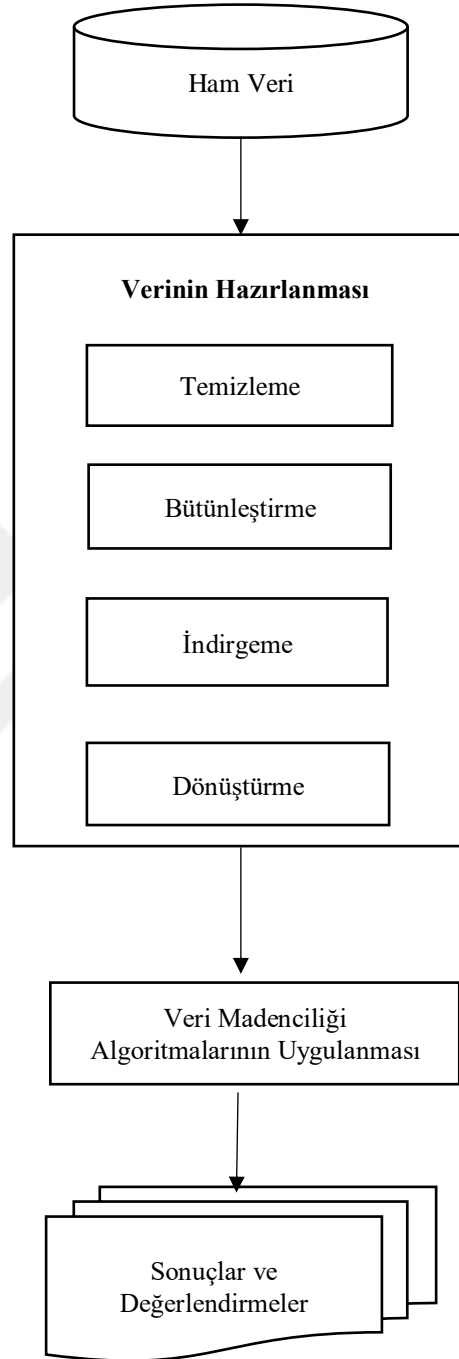
3.2. Veri Madenciliği ve Veri Ön İşleme Teknikleri

Veri madenciliği elde edilen verilerin belirli işlemlerden geçirildikten sonra anlamlı bilgilerin oluşturulmasıdır (MacKinnon ve Glick, 1999). Aynı zamanda, büyük miktarda veri içeren veri setleri üzerinden önemli olanları bulup çıkararak anlamlı bir veri alt kümesi elde etmeye veri madenciliği denir. Veri madenciliği büyük veri grupları içerisinde anlamlı özet veriler çıkarmayı hedefler.

Veri madenciliği algoritmalarının uygulanması öncesinde ham veri üzerinde bazı ön işlemlerin yerine getirilmesi söz konusu olabilir. Kurumların oluşturduğu birçok veri tabanlarında bilgiler eksik, yanlış, tekrarlı ve gereksiz olabilir. Bundan dolayı ham veri birtakım işlemlerden geçtikten sonra veri madenciliği algoritması uygulanarak sonuçlar elde edilmektedir.

Veri tabanları içindeki verinin ve bu veriye dayalı olarak elde edilen veri madenciliği sonuçlarının kalitesinin artırılması, veriyi analize hazırlarken dikkat edilmesi gereken en önemli noktadır. Veri madenciliği işlemlerini kolaylaştırmak ve verimliliği artırmak için veri tabanındaki veriler, bir “ön işleme” aşamasından geçirilir. Ön işleme işlemlerinin uygulanmasının ardından, veri analiz için hazır hale getirilir. Veri madenciliği yöntemlerinin bu işlenmiş veriye uygulanmasıyla sonuçlar elde edilir. Sonuçlar analiz edilerek veri içindeki örüntüler açığa çıkarılır (Çınar ve Dondurmacı, 2014).

Veri hazırlama ile ilgili işlemleri veri madenciliği kavramı içinde düşünürsek, veri madenciliğinin bir süreç olarak değerlendirmesi gerekir. Genellikle veri madenciliği sürecinde yer alan veri ön işlem adımları Şekil 3.1’de verilmiştir. (Han ve Kamber, 2011):



Şekil 3.1. Veri ön işleme adımları

3.2.1. Verinin temizlenmesi

Gerçek uygulamalarda veri eksik, gürültülü veya tutarsız olabilir. Veri temizleme eksik, gürültülü ve tutarsız olan verileri iyileştirmeyi amaçlamaktadır. Eksik veri problemini çözmek için veri tabanından eksik değerlerin bulunduğu kayıt veya kayıtlar atılabilir ya da eksik değerlerin bulunduğu alanlar ilgili değişkenin ortalaması ile doldurulabilir. Verilerde gürültünün kaynağı, bir değer hatalı olarak ölçülmesinden, hatalı veri toplama araçlarından, veri girişi problemlerinden, hatalı nitelik değerlerinden olabilir. Gürültülü veri problemlerini düzeltmek için, demetleme, kümeleme ve eğri uydurma gibi yöntemler kullanılarak veri temizleme işlemi yapılmaktadır.

3.2.2. Verinin bütünleştirilmesi

Birçok farklı kaynaktan gelen ve aynı bilgiyi taşıyan verilerin tek tür veri yapısına dönüştürülüp tek bir çatı altında toplanması işlemine veri bütünleştirilmesi denir. Örneğin cinsiyet veri tipi farklı kaynaklarda değişik şekillerde tutulabilir. Bazı kaynaklarda 0 - 1 şeklinde, bazı kaynaklarda K - E veya Kadın - Erkek, bazı kaynaklarda ise M - F veya Male – Female şeklinde tutulabilir. Bu durumda farklı kaynaklardan gelen bu verilerin aynı tür bilgiyi taşıdıkları için tek tip ve alanda birleştirilmesi gerekmektedir. Veriler arasındaki bu bütünlük yapılan çalışmaların daha doğru sonuç vermesini sağlayacaktır.

3.2.3. Verinin indirgenmesi

Büyük hacimli veri kümelerinden gerekli olmayan nitelikler veya bu niteliğe ait bazı verilerin çıkartılarak daha küçük hacimli veri kümesi elde edilmesi işlemidir. Veri üzerinden faydalı ve doğru sonuç elde etmek için kullanılacak verinin indirgenmesi gerekir. Bu indirgeme işlemi sonucu veri kümesi daha yalın hale gelmektedir. Veri birleştirme, boyut indirgeme, veri sıkıştırma ve kesikli hale getirme yaygın olarak kullanılan veri indirgeme yöntemlerindedir.

3.2.4. Verinin dönüştürülmesi

Farklı kaynaklardan gelen verilerin uygun şekilde analiz edilebilmesi için tek bir veri seti haline getirilmesi işlemi olarak tanımlanır. Veri dönüştürme ile veriler, veri

madenciliği için uygun biçimlere getirilmektedirler. Veri dönüştürme; düzeltme, birleştirme, genelleştirme ve normalleştirme gibi değişik işlemlerden biri veya bir kaçını içerebilir (Roiger ve Geatz, 2003).

3.2.5. Veri madenciliği algoritmalarının uygulanması

Veri ön işleme adımlarından sonra uygulamada kullanılacak veri seti hazır hale gelmiş olur. Veri seti elde edildikten sonra çalışmada hangi algoritmaların kullanılacağına karar verilir. Algoritmalar kullanılarak geçerli sonuçlar elde edilir. Veri madenciliği yöntemleri uygulandıktan sonra elde edilen sonuçlar ilgili kişi ve birimlere sunulur. Sonuçların kullanılan algoritmaya uygun şekil veya grafiklerle desteklenmesi beklenir.

3.2.6. Sonuçlar ve değerlendirmeler

Veri üzerinde veri madenciliği uygulandıktan sonra alınan sonuçlar yorumlanır ve çalışmanın doğru sonuca ulaşp ulaşmadığı araştırılır. Bu aşamada genellikle farklı yöntemler uygulanmışsa onların karşılaştırması yapılır. Elde edilen sonuçlar yapılmış olan diğer çalışmaların sonuçlarıyla karşılaştırılıp doğrulanır.

3.3. Özellik Seçimi Algoritmaları

Özellik seçimi, kullanılan algoritmaya göre özellikleri değerlendirerek veri setindeki n adet özellik arasından en iyi k adet özelliği seçme işlemidir (Forman, 2003). Orijinal veri setini temsil edebilecek en iyi alt kümenin seçimi olarak da tanımlanmaktadır. Bu işlem, ilgilenilen problem için en faydalı ve en önemli özellikleri seçerek veri kümesindeki özellik sayısını azaltmayı yani veri boyutunu düşürmeyi amaçlamaktadır (Ladha ve Deepa, 2011). Veri boyutunun azaltılması kısaca, büyük veri kümelerinin depolanması ve analiz edilmesinde karşılaşılan sorunları aşmak için veri kümesinden ilgisiz veya gereksiz değişkenlerin çıkartılması olarak tanımlanmaktadır.

Bu bölümde çalışmada kullanılan özellik seçim yöntemlerinden Bilgi Kazanımı (BK) ve ReliefF algoritması hakkında bilgi verilmeye çalışılmıştır.

3.3.1. Bilgi kazanımı algoritması

Bilgi Kazanımı (Information Gain), Entropinin (Dağınım, Dağıntı) tersi olarak tanımlanabilir. Entropi, bir sistemdeki belirsizliğin veya tahmin edilemezliğin ölçüsü şeklinde ifade edilir. Entropi ifadesi ilk defa ‘Shannon’ tarafından veri iletişimde ve bilgisayar bilimlerinde kullanılmıştır. Literatürde de Shannon Entropisi (Shannon’s Entropy) olarak da geçen modele göre ‘bir anlatımı ifade etmek için gereken en kısa ihtimallerin ortalama değeri alfabe bulunan tüm sembollerin logaritmasının elde edilen entropiye bölümüdür’ (Şeker, 2008) şeklinde bir kuram ortaya atmıştır. Yani ifade edilecek mesajdaki değişim ne kadar fazla olacak ise o kadar fazla kodlamaya ihtiyaç duyulmaktadır.

Bilgi kazanımı entropinin tersidir ve $[0,1]$ aralığında ifade edilir. Verilen bir niteliğin elde edilen sınıflandırma sonuçlarını ne kadarlık bir değer ile etkileyebileceğini gösterir. Her farklı sınıf için farklı farklı değerler alan bir nitelik varsa entropi 0 çıkacak ve bilgi kazanımı 1 olacaktır. Bu ifade seçilen nitelik ile sınıfın arasında birebir bir bağlantı olduğudur. Nitelik sınıfa ne kadar bağlıysa bilgi kazanımı o kadar yüksek, ne kadar bağımsızsa bilgi kazanımı o kadar düşük çıkar (Tunç, 2016). Denklem 3.1, 3.2 ve 3.3’te hesaplama denklemleri verilmiştir (Hall, 1999).

$$Bilgi(D) = - \sum_{i=1}^m (p_i \log_2 p_i) \quad (3.1)$$

$$Bilgi_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} x Bilgi(D_j) \quad (3.2)$$

$$Bilgi Kazancı = Bilgi(D) - Bilgi_A(D) \quad (3.3)$$

Denklem 3.1, 3.2 ve 3.3’te yer alan değerlerin açıklamaları aşağıdaki gibidir:

- m : hedef değişkenin alabileceği değerlerin adedi
- A : tahmin edici değişken
- D : hedef değişken
- v : tahmin edici değişkenin alabileceği değerlerin adedi

Yöntemin zayıf yanı, daha fazla bilgiye sahip olmasa bile çok çeşitli değerlere sahip özellikler lehine önyargılı şekilde sonuç vermesidir (Novakovic ve ark., 2011).

3.3.2. ReliefF algoritması

Relief, Kira ve Rendell tarafından 1992'de önerilen ikili sınıflandırmada kullanılan bir özellik seçimi algoritmasıdır. İki sınıfı olan veri setleri için çalışan bu algoritma veri setleri çoklu sınıfa sahip olduğu zaman çalışmamaktadır. Kononenko, bu çok sınıflı problemlere çözüm üretmek için Relief algoritmasına bazı güncellemeler önermiştir ve bu algoritma ReliefF olarak adlandırılmıştır.

ReliefF algoritmasının temel çalışma prensibi, veri seti içerisindeki her bir özelliğin benzer yapıdaki görüntü nesnelere ayırt edebilme kabiliyetinin ölçülmesi esasına dayanmaktadır. Bu noktada “Belirli bir mesafeye yakın uzaklıkların aynı sınıfa ait olma olasılığı yüksektir” prensibi göz önüne alınmaktadır (Kononenko, 1994).

ReliefF, Relief istatistiksel modelinin geliştirilmiş versiyonudur. ReliefF metodu, veri setinden bir örnek ele alarak ilgili örneğin, kendi sınıflarındaki diğer örneklerle yakınlığını ve farklı sınıflarla olan uzaklığına bağlı bir model oluşturarak öznelik seçme işlemini gerçekleştirmektedir.

Denklem 3.4'te, ReliefF hesaplama denklemi verilmiştir (Bolón-Canedo ve ark. 2014).

$$S_i = \frac{\sum_{j=1}^m -fark(x_{ij}, enyakın_aynı_{ij}) + fark(x_{ij}, enyakın_farklı_{ij})}{m} \quad (3.4)$$

Denklem 3.4'te yer alan değerlerin açıklamaları aşağıdaki gibidir:

m : veri setinde bulunan örnek sayısı

S_i : i 'ninci niteliğin ReliefF değeri

$fark(x_{ij}, enyakın_aynı_{ij})$: j 'ninci örnekte bulunan i 'ninci niteliğin aynı sınıfa sahip en yakın örneğe olan uzaklığı

$fark(x_{ij}, enyakın_farklı_{ij})$: j 'ninci örnekte bulunan i 'ninci niteliğin farklı sınıfa sahip en yakın örneğe olan uzaklığı

3.4. Sınıflandırma Algoritmaları

Sınıflandırma kavramı, bir veri kümesinde bulunan değerlerin sınıflara ayrılmasıdır. Sınıflandırma problemi için geliştirilen algoritmalar, adından da anlaşılacağı üzere verileri belli özelliklerine göre sınıflandırır. Eğer sistem, hangi verinin, hangi koşullarda, hangi sınıfa ait olacağı bilgisi ile sınıflandırılarak eğitilirse, yeni veri setindeki veriyi de öğrendiklerine benzer biçimde sınıflandırabilir.

Bir sınıflandırma işleminde kullanılan veri temel olarak ikiye ayrılır; verilerin ilk bölümü eğitim verisi diğeri ise test verisidir. Sınıflandırma modeli eğitim verisi ile eğitilir ve test verisi ile de değerlendirilir. Test verilerinde elde edilen sınıflandırma başarısı modelin doğruluğunu verir.

Eğitim seti sınıflandırma modelinin elde edileceği kayıtları içerir. Eğitim setindeki her kayıt bir tanesi sınıf özneliği olan özneliklerden oluşur. Test seti ise bir sınıflandırma modelinin doğruluğunu belirlemek için kullanılır. Çalışmalarda sık karşılaşılan sınıflandırma yöntemlerinden bazıları şunlardır:

- Lojistik regresyon
- Yapay sinir ağları
- Naif bayes
- Destek vektör makineleri
- Karar ağaçları
- k-en yakın komşuluk

Bu bölümde çalışmada kullanılan LR, YSA ve DVM algoritmaları incelenmektedir.

3.4.1. Lojistik regresyon

LRA bağımlı değişken ile bir veya birden fazla bağımsız değişken arasındaki ilişkiyi modellemek için kullanılan istatistiksel bir yöntemdir. Burada bağımlı değişken kategorik, bağımsız değişken ise sürekli veya kategorik düzeyde olabilmektedir (Hosmer ve Lemeshow, 2000). Bağımlı değişkenin sadece iki kategoriye sahip olduğu modeller ikili lojistik regresyon analizi olarak bilinmektedir (Bayram, 2017). Yapılan çalışmalarda bağımlı değişken genellikle sadece iki sonuca sahiptir.

LR'nin sonuç değişkeninin ikili veya çoklu olması onu doğrusal regresyondan ayıran en önemli özelliktir. Lojistik regresyon ve doğrusal regresyon analizinin ortak özelliği her ikisinde de bazı değişken değerlerine dayanarak tahmin yapılmak istenmesidir (Elhan, 1997).

LR ikili kategorik bağımlı değişkene sahiptir. Var/yok, kazan/kaybet, kredi ver/kredi verme, hasta/hasta değil vb. yani bağımlı değişkenimiz 1 veya 0 değerlerinden birisini alır. Regresyon formülüne baktığımızda olasılık hesaplarının modelin temelini oluşturduğunu görebiliriz. LR modeli denklem 3.5'teki gibi formüle edilmiştir.

$$L = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_i X_i + \varepsilon_i \quad (3.5)$$

Denklem 3.5'te yer alan değerlerin açıklamaları aşağıdaki gibidir. (Hosmer, Lemeshow, 2000):

- L : logit fonksiyonu
- P_i : tahmin edilen olasılık
- β_0 : sabit katsayı
- β_i : bağımsız değişken katsayısı
- X_i : bağımsız değişken
- ε_i : hata terimi

$P(x)$ olasılığı formülü denklem 3.6'te verilmiştir.

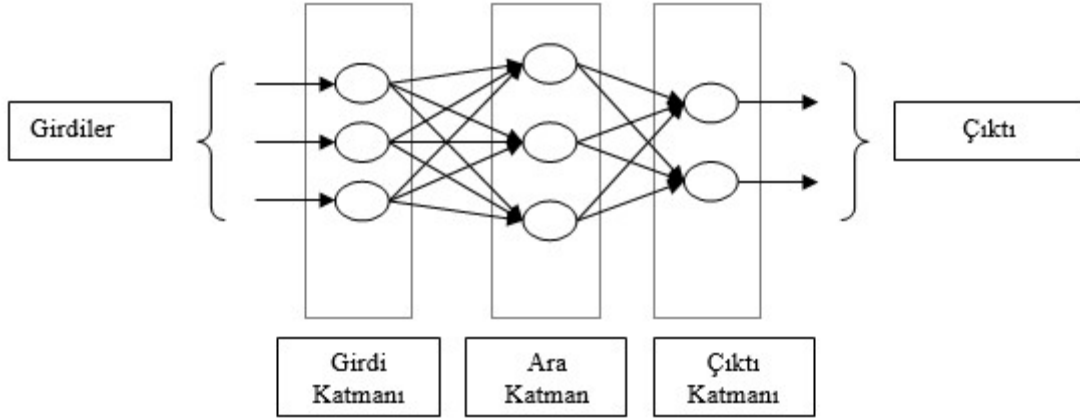
$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}} \quad (3.6)$$

3.4.2. Yapay sinir ağları

YSA insan beyninden esinlenerek geliştirilmiş ve biyolojik sinir ağlarına benzer yapıda özellikleri içeren bilgi işleme yapıları olarak ifade edilmektedir (Fausett, 1994).

Nöronların bir araya gelmeleri genel bir YSA sisteminde katmanları oluşturur. Şekil 3.4'te görüldüğü üzere bir YSA'da, girdi katmanı (input layer), ara katman (hidden layer) ve çıktı katmanı (output layer) olmak üzere üç temel katman bulunmaktadır. Girdi katmanı, dışarıdan girilen girdilerin ağırlıklandırılarak ara katmana iletilmesini sağlayan ilk katmandır. Bu girdilerin istatistikteki karşılığı bağımsız değişkenlerdir. Ara katman girdi ve çıktı katmanları arasında kalan kısımdır. Bu katmanın dış ortamla doğrudan bir

bağlantısı bulunmaz. Sadece girdi katmanından gönderilen sinyalleri alarak çıktı katmanına iletirler. Çıktı katmanı sinir ağındaki son katmandır. Bilgilerin dış ortama aktarılmasını sağlar. Çıktı değişkenlerinin istatistikteki karşılığı bağımlı değişkenlerdir.



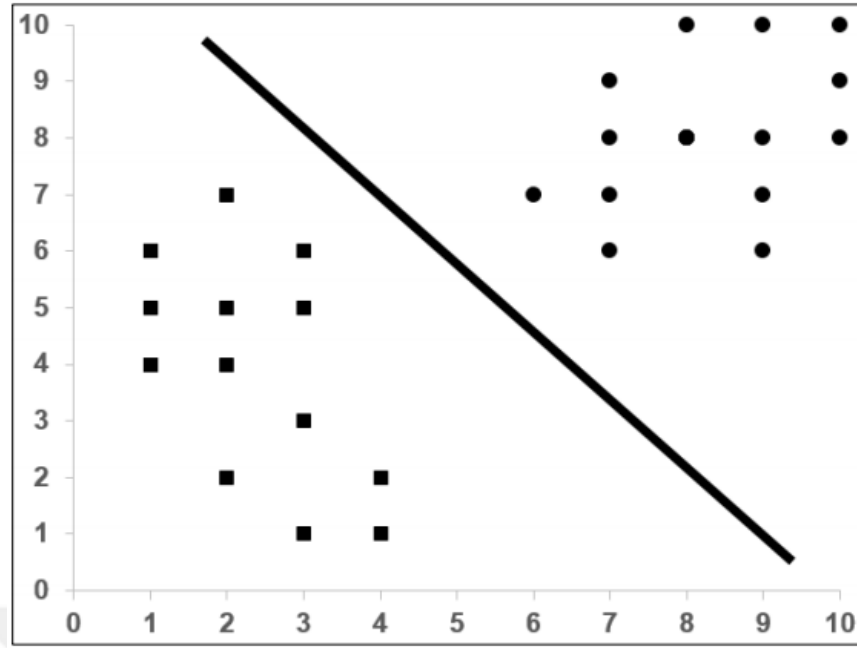
Şekil 3.2. YSA örneği

Genel olarak YSA, insan beyninde olduğu gibi verinin eğitilmesi, öğrenilmesi, genelleme yapabilmesi ve çok sayıda değişkenle çalışabilmesini amaçlamaktadır (İslamoğlu, 2015).

3.4.3. Destek vektör makineleri

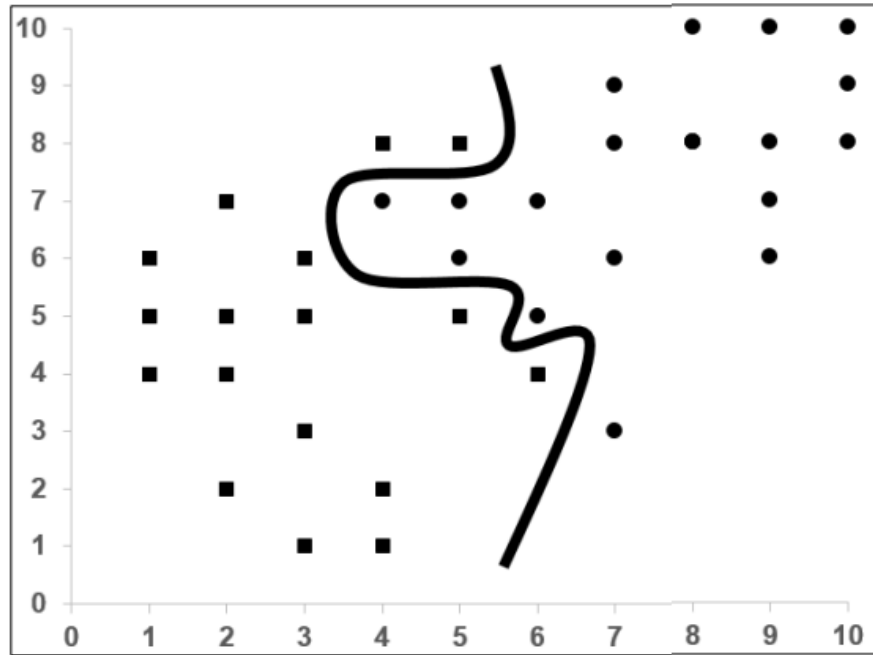
DVM, regresyon problemlerinin yanı sıra genellikle sınıflandırma problemleri için kullanılan oldukça basit, etkili ve gözetimli (supervised) bir makine öğrenmesi algoritmasıdır. DVM, sınıflandırma ve regresyon tipindeki problemlerin çözümü için Vapnik tarafından geliştirilmiştir. Temel amacı sınıfları birbirinden ayıracak optimal hiper düzlemin elde edilmesi prensibine dayanmaktadır (Vapnik, 1995). DVM, doğrusal ve doğrusal olmayan destek vektör makineleri olmak üzere iki durumda değerlendirilir.

- Doğrusal DVM: Eğitim için kullanılacak N elemandan oluşan verinin $\theta = \{x_i, y_i\}$, $i = 1, 2, \dots, N$ olduğu varsayalım. Burada $y_i \in \{-1, 1\}$ etiket değerleri ve $x_i \in R^d$ özellikler vektörüdür. Doğrusal DVM'de veriler direkt olarak bir düzlem ile ayrılabilir. DVM'lerin amacı bu düzlemin iki ayrı sınıfta bulunan örnek grubuna eş uzaklıkta olmasını sağlamaktır (Yakut ve ark., 2014). Şekil 3.3'de doğrusal DVM gösterilmiştir.



Şekil 3.3. Doğrusal DVM

- Doğrusal Olmayan DVM: Doğrusal olarak ayrılan DVM'lerde veriler iki sınıfa doğrusal bir düzlem ile ayrılabilirdi. Uygulamada bu durum her zaman geçerli olmayabilir. Doğrusal bir düzlem ile ayrılmayan verilerin gösterimi Şekil 3.4'te gösterilmiştir (Yakut ve ark., 2014).



Şekil 3.4. Doğrusal olmayan DVM

DVM'in önemli avantajlarından birisi sınıflandırma problemini kareli optimizasyona dönüştürerek çözmesidir. Bu sayede öğrenme aşamasındaki işlem sayısı azaltılarak daha hızlı performans sergilenmektedir. Bu durum büyük veri setlerinde avantaj sağlamaktadır (Nitze ve ark., 2012).

3.5. K-kat Çapraz Doğrulama

K-kat çapraz doğrulama, veri setini parçalara ayırarak modellerin eğitilmesini sağlayan bir yöntemdir. Bu yöntemde veri seti k eşit parçaya bölünür. k-1 parça eğitim için kullanılırken 1 parça da test seti için kullanılır ve k defa bu işlem tekrar eder. Performans ölçülerine göre en iyi durum ölçülerin ortalaması alınarak belirlenir. Bu sayede her veri eğitim ve test aşamalarında kullanılmış olduğu için sapmaların ve hataların en az seviyeye inmesi sağlanmaktadır (Şirin, 2017).

K değerinin en sık kullanılan değeri 10 dur. k büyük seçilirse, tahmin daha doğru olacağından gerçek hata tahmininin doğruluktan sapması küçük, varyansı ve hesaplama zamanı büyük olur. k sayısı küçük seçilirse tahminin doğruluğundaki değişim küçük ve doğruluktan sapması gerçek hata tahminine göre daha büyük olur.

Şekil 3.5'te K-kat çapraz doğrulama sisteminin k değerinin 10 seçildiği örnek gösterilmektedir. Koyu renk kutucuklar test veri kümesini açık renk kutucuklar eğitim veri kümesini göstermektedir.

	1. Parça	2. Parça	3. Parça	4. Parça	5. Parça	6. Parça	7. Parça	8. Parça	9. Parça	10. Parça
1. Adım	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
2. Adım	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
3. Adım	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
4. Adım	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
5. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
6. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim
7. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim
8. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim
9. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim
10. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test

Şekil 3.5. K-10 çapraz doğrulama yöntemi

3.6. Karmaşıklık Matrisi

Test sonucu elde edilen bilgiler üzerinden, gerçek ve tahmin edilen örneklerin sayılarını temel alan bir matris ile algoritmanın başarısını ölçmeye yarar bir yöntemdir (Powers, 2011). Karmaşıklık matrisinde satırlar test kümesindeki gerçek sınıfa ait değerleri, kolonlar ise tahmin edilen sınıfa ait değerleri ifade eder. Şekil 3.6'da matrisin içeriği gösterilmektedir.

		Tahmin Edilen Sınıf	
		Sınıf = 1	Sınıf = 0
Gerçek Sınıf	Sınıf = 1	DP	YN
	Sınıf = 0	YP	DN

Şekil 3.6. Karmaşıklık matrisi

Karmaşıklık matrisi için elde edilen değerlerin açıklamaları aşağıdaki gibidir.

- Doğru Pozitif (DP): Gerçek değeri 1 olup tahmin edilen değer de 1 olduğu kayıtlardır.
- Doğru Negatif (DN): Gerçek değeri 0 olup tahmin edilen değer de 0 olduğu kayıtlardır.
- Yanlış Pozitif (YP): Gerçek değeri 0 fakat tahmin edilen değer 1 olduğu kayıtlardır.
- Yanlış Negatif (YN): Gerçek değeri 1 fakat tahmin edilen değer 0 olduğu kayıtlardır.

Doğruluk oranı, kesinlik, duyarlılık ve F-ölçütü sınıflandırma başarımını değerlendirmede kullanılan temel ölçütlerdir. Modelin başarısı, doğru ve yanlış sınıfa atanan örneklerin sayıları ile ilişkilidir.

Doğruluk: Sınıflandırma performansını değerlendirmek için kullanılan en yaygın yöntemdir. Doğru sınıflandırılan örnek sayısının toplam örnek sayısına oranıdır.

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (3.7)$$

Kesinlik: Gerçek değeri pozitif olup pozitif değere sınıflandırılan sayısının, pozitif değere sınıflandırılanların toplamına oranıdır.

$$\text{Kesinlik} = \frac{DP}{DP+YP} \quad (3.8)$$

Duyarlılık: Doğru sınıflandırılan pozitif örnek sayısının toplam pozitif örnek sayısına oranıdır.

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (3.9)$$

F-ölçütü: Kesinlik ve duyarlılığın harmonik ortalamasıdır.

$$F - \text{Ölçütü} = \frac{2 * \text{Duyarlılık} * \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (3.10)$$

ROC Eğrisi: Testlerin doğru karar verme gücünü değerlendirmede kullanılan ölçütlerden biri ROC (Receiver Operating Characteristics) eğrisi ve altında kalan alandır. Eğri altında kalan alan “Area Under the Curve (AUC)” olarak tanımlanır ve en büyük “1” değerini alabilir. AUC değeri 0 ile 1 arasında değer almak ile birlikte 1 değerine yakınsaması, yapılan testin başarısındaki artışı göstermektedir (Davis ve Goadrich, 2006).

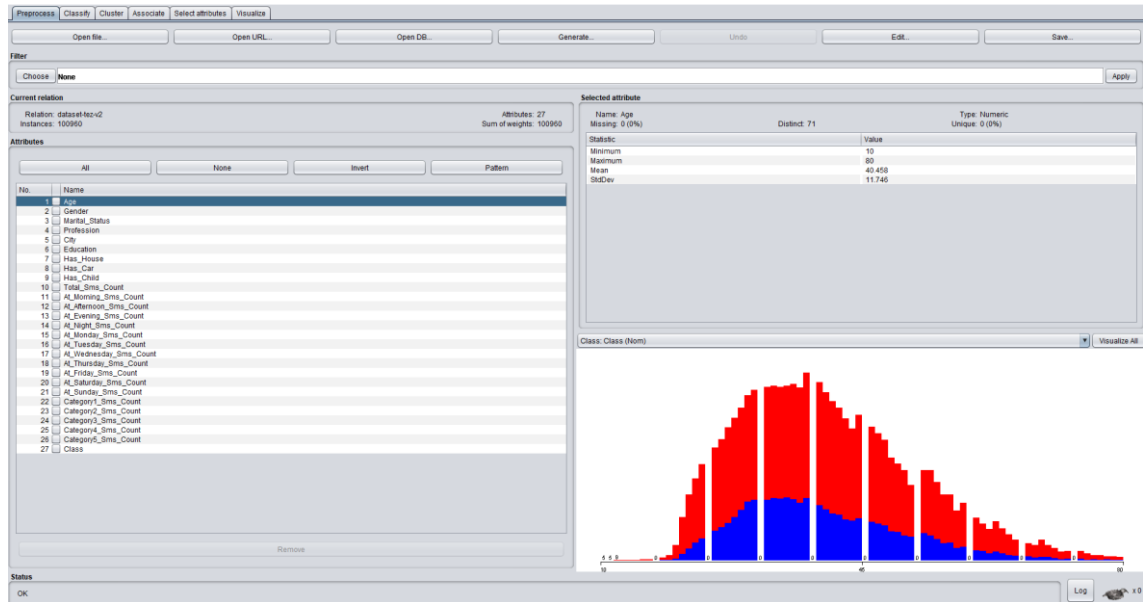
3.7. Kullanılan Yazılım Geliştirme Ortamları

Veri madenciliğinin amacı ham veriyi anlamlı, etkin ve yararlı olan bilgiye dönüştürebilmektir. Bu amaca ulaşabilmek için de veri madenciliği konusunda geliştirilmiş yazılımların kullanılması veri madenciliği süreçlerini kolaylaştırmaktadır. Bu alanda sıklıkla kullanılan yazılımlar şunlardır: Keel, Knime, Orange, RapidMiner, Weka, Microsoft Azure ML Studio.

Bu bölümde çalışmada kullanılan Weka ve Azure ML Studio geliştirme ortamları hakkında bilgi verilmeye çalışılmıştır.

3.7.1. WEKA

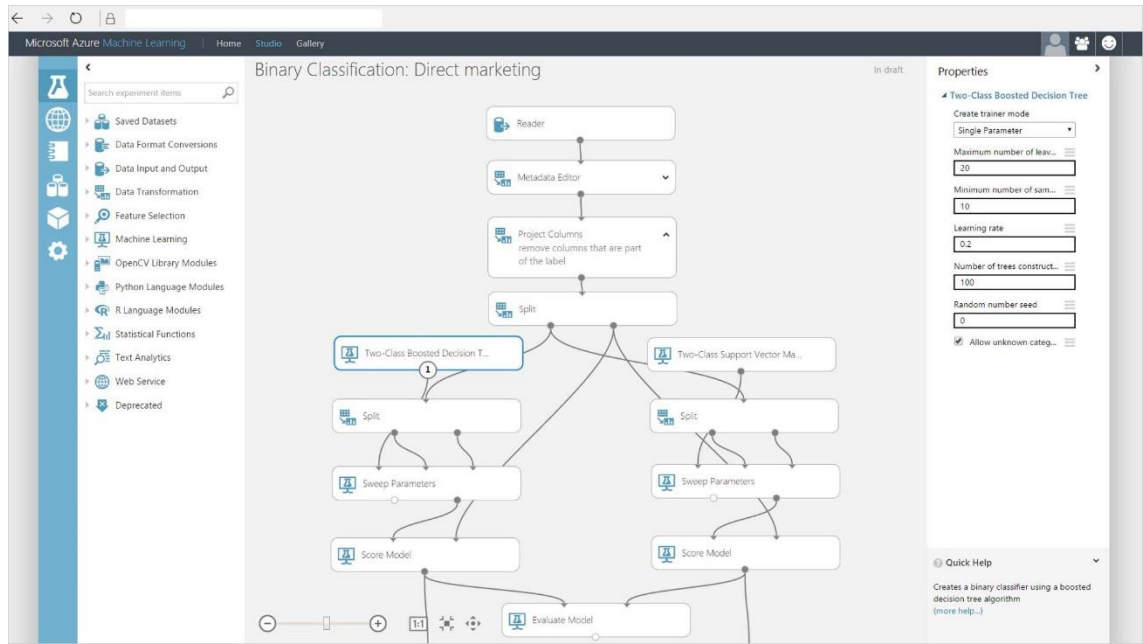
WEKA, Waikato Üniversitesinde, Java platformu üzerinde geliştirilmiş ve GNU lisansı altında olan açık kaynak kodlu bir veri madenciliği programıdır. Veri ön işleme, özellik seçimi, sınıflandırma, kümeleme gibi çeşitli veri madenciliği işlemleri Weka ile yapılabilmektedir. Arff (Attribute Relationship File Format) dosya formatı Weka için tasarlanan bir dosya formatıdır (Witten ve Frank, 2005).



Şekil 3.7: Weka programına ait bir ekran görüntüsü

3.7.2. Azure machine learning studio

Microsoft Azure Machine Learning Studio, veriler için tahmine dayalı analiz çözümleri oluşturma, test etme ve dağıtma amacıyla kullanılabilen işbirliğine dayalı sürükle ve bırak aracıdır. Tahmine dayalı bir analiz modeli geliştirmek için, genellikle bir veya daha çok kaynaktan veri kullanır, çeşitli veri işleme ve istatistik işlevleri aracılığıyla bu verileri dönüştürüp analiz eder ve bir sonuç kümesi oluşturur.



Şekil 3.8: Azure ML Studio programına ait bir ekran görüntüsü

Azure Machine Learning Studio, tahmine dayalı bir analiz modelini kolayca oluşturma, test etme ve yineleme amacıyla etkileşimli ve görsel bir çalışma alanı sunar. Veri kümelerini ve analiz modüllerini tuvale sürükleyip bırakarak ve bunları birbirine etkileşimli bir şekilde bağlayarak bir deneme oluşturma imkanı sağlar. Aynı zamanda, modelleri özel uygulamalar veya Excel gibi Business Intelligence (BI) araçları tarafından kolayca kullanılabilen web hizmeti olarak yayımlar (Microsoft Dokümantasyonu).

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Bu bölümde tez boyunca yapılmış olan çalışmalar ve elde edilen sonuçlar paylaşılmıştır.

4.1. Kullanılan Veri Seti

Çalışmada özel bir bankanın 2016-2017 yılları arasında mobil pazarlama amacıyla göndermiş olduğu veriler kullanılmıştır. Veri seti; 100.960 satır ve 27 sütunun yer aldığı iki boyutlu bir matristir. Sütunlar müşterilere ait 26 nitelik ve bir sınıf verisinden oluşmaktadır. Sınıf verisi müşterilerin pazarlama SMS'lerini engelleyip engellemediğine dair sınıf bilgisidir. Her bir satır bir müşteriye ifade etmektedir. Nitelikler, müşterinin bazı demografik bilgileri ve müşteriye gönderilmiş 1.755.513 adet pazarlama SMS'leriyle ilgili istatistiki bilgilerden oluşmaktadır. Veri seti içerisinde yer alan özellikler ve açıklamaları Çizelge 4.1'de belirtilmiştir.

Çizelge 4.1. Veri seti içerisinde yer alan nitelikler ve açıklamaları

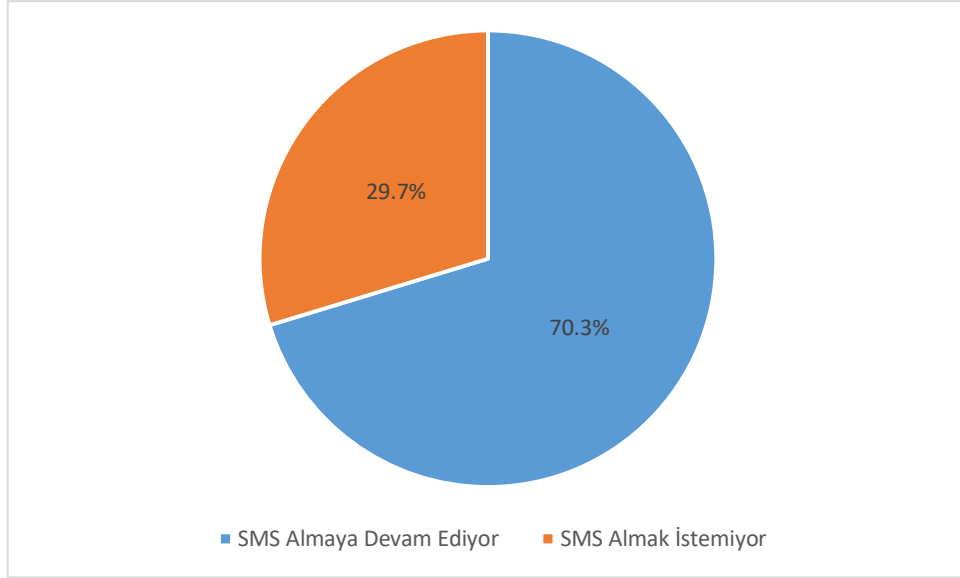
Nitelik	Açıklama
Age	Yaş
Gender	Cinsiyet
Marital_Status	Evlilik Durumu
Profession	Meslek
City	Yaşanılan Şehir
Education	Eğitim Durumu
Has_House	Evi Var Mı
Has_Car	Arabası Var Mı
Has_Child	Çocuğu Var Mı
Total_Sms_Count	Toplam Gönderilen SMS Sayısı (GSS)
At_Morning_Sms_Count	Öğleden Önce GSS
At_Afternoon_Sms_Count	Öğleden Sonra GSS
At_Evening_Sms_Count	Akşam GSS
At_Night_Sms_Count	Gece GSS
At_Monday_Sms_Count	Pazartesi GSS
At_Tuesday_Sms_Count	Salı GSS
At_Wednesday_Sms_Count	Çarşamba GSS
At_Thursday_Sms_Count	Perşembe GSS
At_Friday_Sms_Count	Cuma GSS
At_Saturday_Sms_Count	Cumartesi GSS
At_Sunday_Sms_Count	Pazar GSS
Category1_Sms_Count	Kategori-1: Altın, Döviz, Yatırım İşlemleri Tipinde GSS
Category2_Sms_Count	Kategori-2: Alışveriş, Ulaşım, Eğitim, Giyim Tipinde GSS
Category3_Sms_Count	Kategori-3: Hesap İşlemi, Mobil, Kredi Kullanımı Tipinde GSS
Category4_Sms_Count	Kategori-4: Kredi Kartı, Ödemeler Tipinde GSS
Category5_Sms_Count	Kategori-5: Bilgilendirme ve Bilinmeyen Tipinde GSS
Class	SMS Almaya Devam Ediyor Mu

Veri setinde yer alan nitelikler için ortalama, minimum (Min) deęer, maksimum (Maks) deęer, standart sapma ve farklı deęer sayısı verileri izelge 4.2’de verilmiřtir.

izelge 4.2. Veri setine ait bazı deęerler

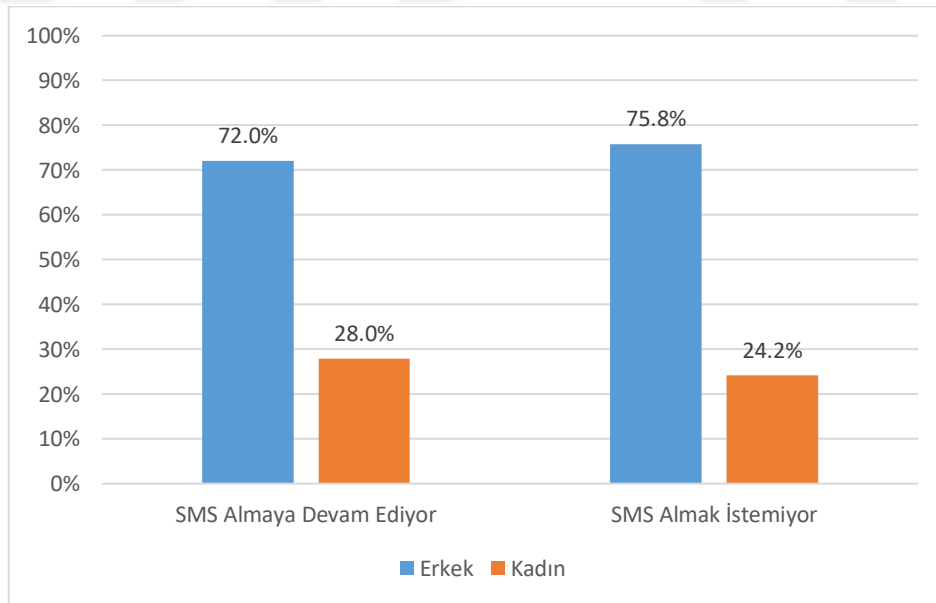
Nitelik	Min	Maks	Ortalama	Standart Sapma	Farklı Deęer Sayısı
Yař	10	80	40.458	11.746	71
Cinsiyet	0	1	0.268	0.443	2
Evlilik Durumu	0	1	0.272	0.445	2
Meslek	0	61	23.12	11.73	62
Yařanılan Őehir	1	190	85.403	45.658	136
Eęitim Durumu	1	4	2.879	0.791	4
Evi Var Mı	0	1	0.493	0.5	2
Arabası Var Mı	0	1	0.455	0.498	2
ocuęu Var Mı	0	1	0.477	0.499	2
Toplam GSS	5	70	17.388	10.391	66
Öęleden Önce GSS	0	27	6.442	3.754	28
Öęleden Sonra GSS	0	37	7.867	5.061	38
Akřam GSS	0	22	3.003	2.914	23
Gece GSS	0	4	0.077	0.278	5
Pazartesi GSS	0	17	2.105	1.997	18
Salı GSS	0	25	4.751	2.991	26
arřamba GSS	0	18	3.086	2.199	19
Perřembe GSS	0	15	2.797	2.018	16
Cuma GSS	0	25	4.497	3.609	26
Cumartesi GSS	0	1	0.042	0.2	2
Pazar GSS	0	2	0.111	0.315	3
Kategori-1	0	31	5.628	4.317	31
Kategori-2	0	27	3.08	3.774	26
Kategori-3	0	21	3.284	2.126	22
Kategori-4	0	13	1.785	1.818	13
Kategori-5	0	26	3.611	3.064	23

Veri seti ierisinde yer alan sınıf bilgisi daęılımını Őekil 4.1’de verilmiřtir. Veri setinde yer alan müşterilerin %29.7’si pazarlama SMS’leri almak istemezken, %70.3’ü pazarlama SMS’i almaya devam etmektedir.



Şekil 4.1. Veri seti sınıf dağılımı

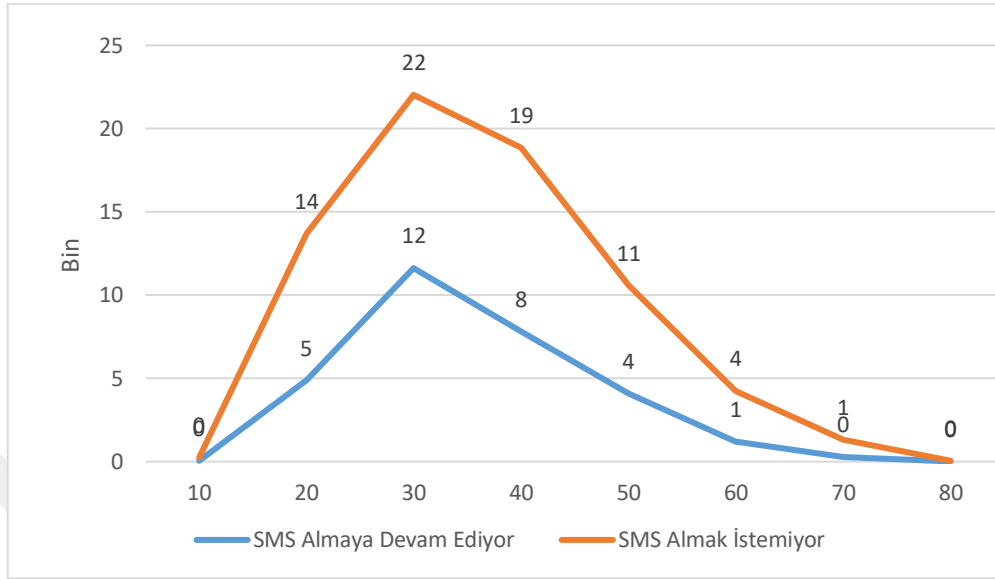
SMS almaya devam eden ve SMS almak istemeyen müşterilerin cinsiyete göre dağılımları Şekil 4.2’te gösterilmiştir. İki sınıf içinde erkek müşterilerin veri setinde ağırlıklı olduğu gözlenmektedir. SMS almaya devam eden müşterilerdeki kadın oranı almak istemeyen müşterilerdekinden azda olsa fazla olduğu gözükmektedir.



Şekil 4.2. Cinsiyete göre sınıf dağılımı

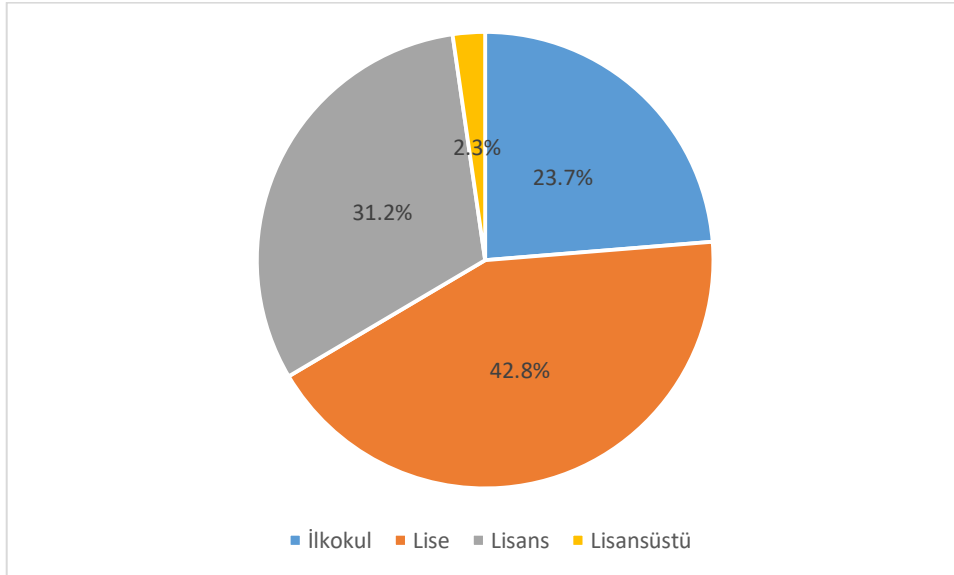
Veri seti içerisindeki müşterilerin yaş dağılımları Şekil 4.3’da gösterilmiştir. Veri setinde bulunan müşteriler ağırlıklı olarak 20-40 yaş aralığındadır. SMS almaya devam

eden ve almak istemeyen müşterilerde yaş yükseldikçe veri seti içerisindeki oranı azalmaktadır. Veri seti içerisinde en çok 30'lu yaşlardaki müşteriler bulunmaktadır.



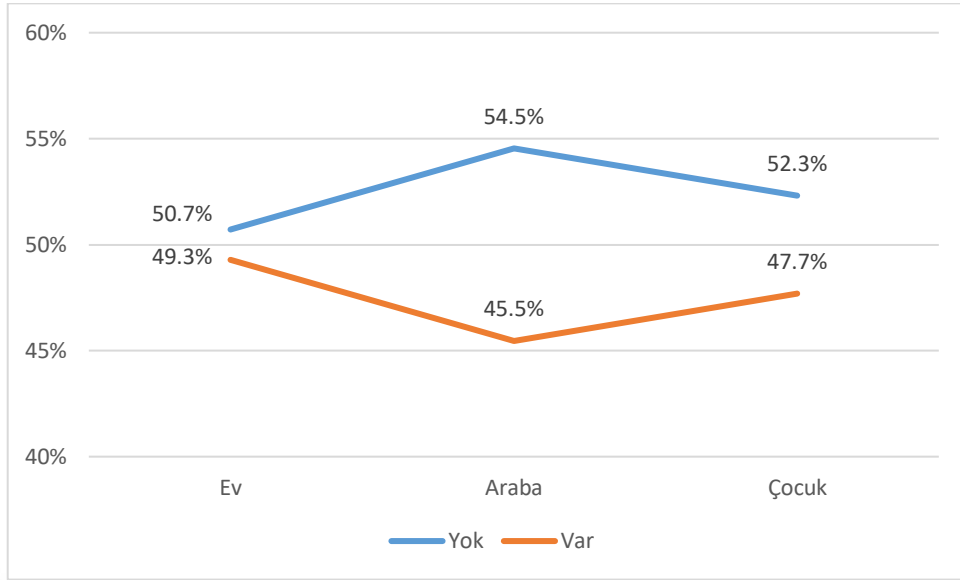
Şekil 4.3. Yaşa göre sınıf dağılımı

Müşteriler en az ilkokul seviyesinde eğitim düzeyine sahiptir. Müşterilerin büyük bir kısmı lise eğitim düzeyine sahiptir.



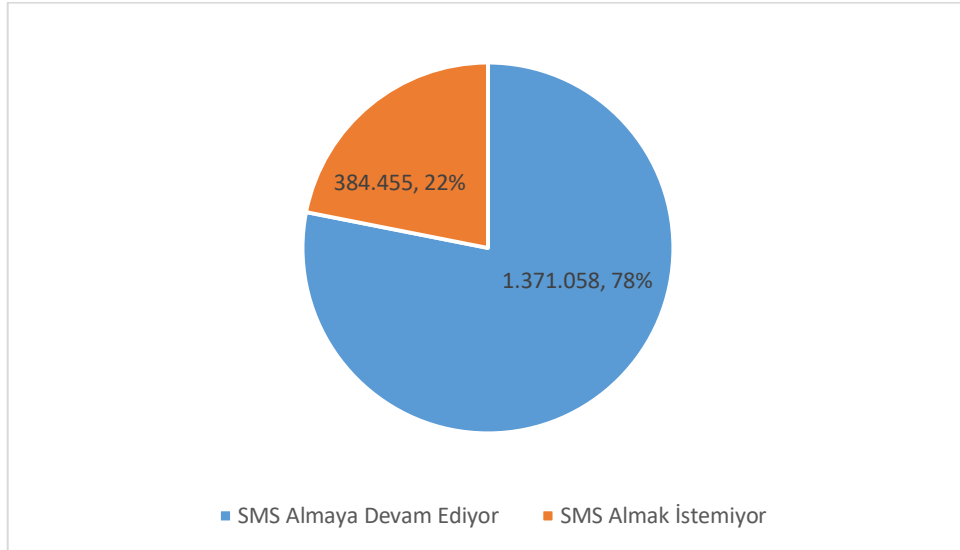
Şekil 4.4. Eğitim seviyesi dağılımı

En az bir eve sahip olan ve olmayan müşteri oranı birbirlerine yakın olarak gözlenmektedir. Arabası olmayan müşteriler veri seti içerisinde ağırlıklı olarak gözükmemektedir. Çocuğu olmayan müşteriler çocuk sahibi olan müşterilerden fazladır.



Şekil 4.5. Ev, araba, çocuk durumu

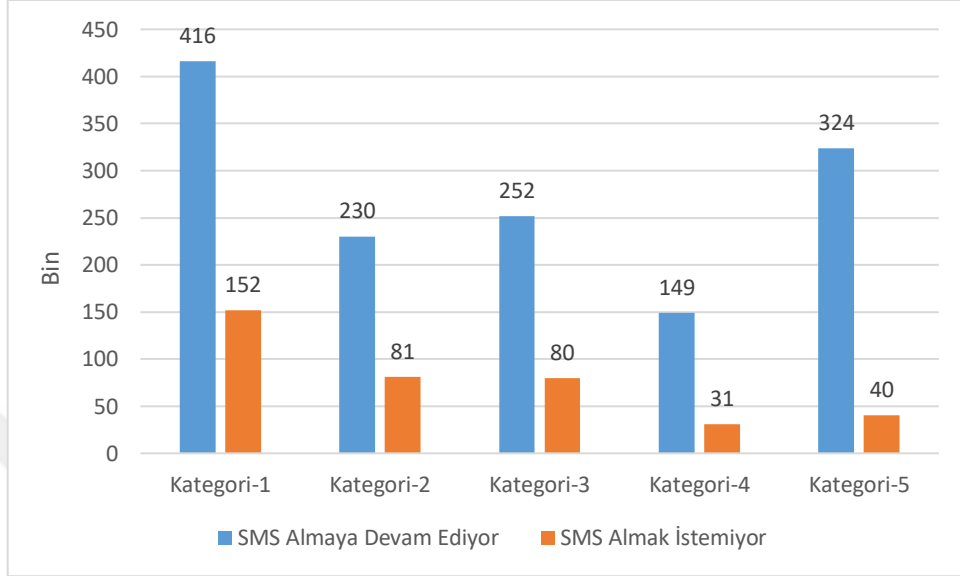
Veri setinin sınıflara göre SMS sayısı dağılımı Şekil 4.6'da verilmiştir. 1.755.513 SMS verisinden oluşturulan veri setinde SMS almak isteyen müşterilere gönderilen SMS sayısı 1.371.058 iken SMS almak istemeyenlere gönderilen SMS sayısı 384.455 olarak gözükmektedir.



Şekil 4.6. SMS sayıları dağılımı

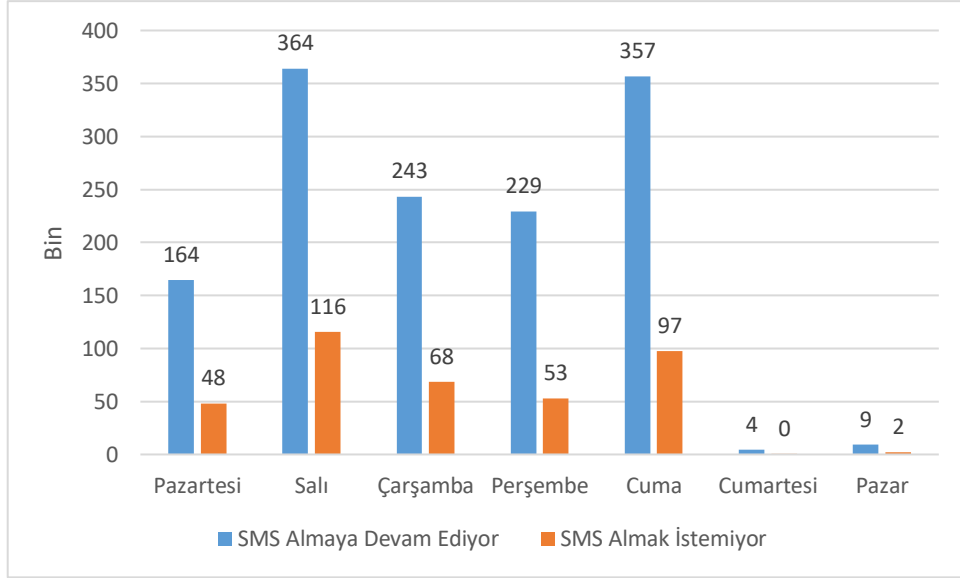
Pazarlama amacıyla gönderilen SMS'ler içeriklerine göre kategorilere ayrılmıştır. Kategori-1 içeriğinde gönderilen SMS sayısı hem SMS almaya devam eden hem de almak

istemeyen müşterilerde diğer kategorilerden oldukça fazla olduğu gözükmektedir. SMS almaya devam eden ve almak istemeyen müşterilere kategori-5 içerikte gönderilen SMS sayılarındaki fark oldukça fazladır.



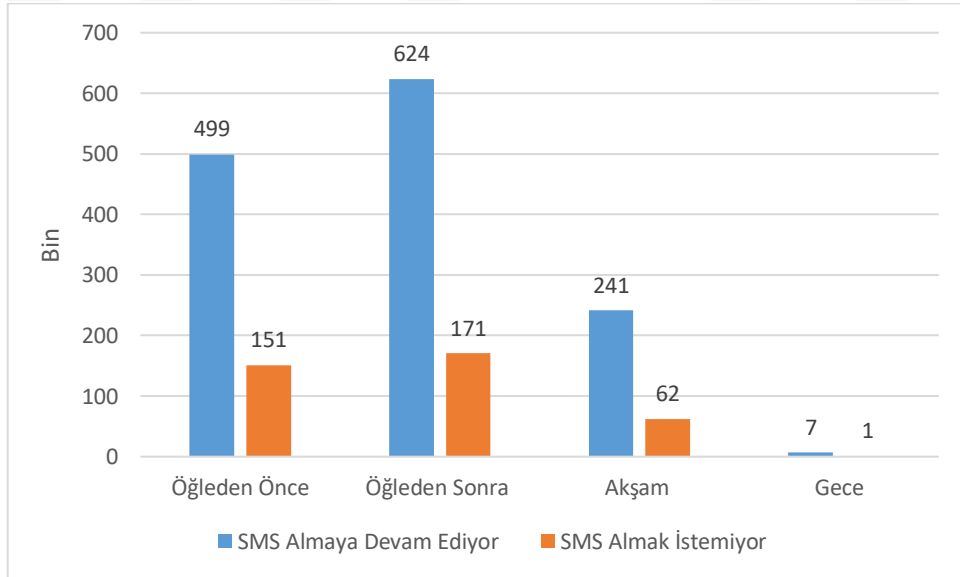
Şekil 4.7. Kategoriye göre SMS sayıları

Şekil 4.8’de günlere göre gönderilen SMS sayıları verilmiştir. SMS almaya devam eden ve istemeyen müşterilere gönderilen SMS sayılarının en fazla olduğu gün Salı günüdür. Cuma gününde SMS almaya devam eden müşterilere gönderilen SMS’lerde ciddi artış olduğu gözlenmektedir. Hafta sonu gönderilen SMS sayıları oldukça azdır. Hafta sonu gönderilen SMS sayısının az olması banka tarafından alınmış bir karar olduğu düşünülmektedir. Tüm günlerde SMS almaya devam eden müşterilere gönderilen SMS sayısı almak istemeyenlere göre fazladır.



Şekil 4.8. Haftanın günlerine göre SMS sayıları

Gün içi zaman aralığında gönderilen SMS sayılarında gündüz saatlerinde gönderilen SMS'ler ağırlıktadır. Gece giden SMS'ler çok az miktardadır. Gece SMS gönderilmemesi müşteri memnuniyeti için alınmış bir karar olduğu düşünülmektedir. Gündüz gönderilen SMS'ler öğleden sonra ağırlıklı olarak gönderilmiştir.



Şekil 4.9. Gün içi zaman aralıklarına göre SMS sayıları

4.2. Özellik Seçimi

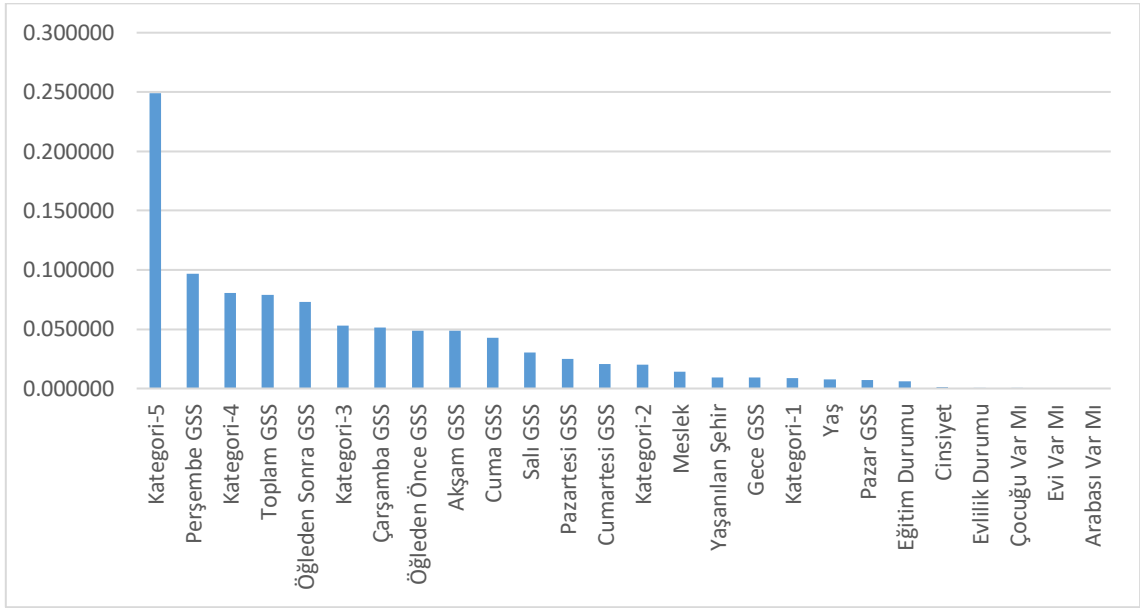
Özellik seçimi işlemi, sınıflandırma başarımını artırmak ve model oluşturma sürelerini azaltmak amacıyla özellik alt kümesi seçilmesi işlemidir. (Awad ve ark., 2013). Çalışma kapsamında özellik seçimi işlemleri WEKA uygulaması üzerinde yapılmıştır. Özellik seçimi işlemleri Windows 10 64-bit işletim sistemi üzerinde Intel i7-4700MQ @2,40Ghz CPU ve 8.0 GB RAM özellikli kişisel bilgisayarda yapılmıştır.

4.2.1. Bilgi kazanımı özellik seçimi sonuçları

Bilgi Kazanımı özellik seçimi yöntemi uygulandığında elde edilen sonuçlar önem derecelerine göre Çizelge 4.3 ve Şekil 4.10'da verilmiştir. Elde edilen sonuçlar incelendiğinde kategori-5 niteliğinin önem derecesinin diğer niteliklere göre oldukça fazla olduğu gözükmektedir. Kategori-5 içeriğinde gönderilen SMS sayılarının veri seti içerisinde oldukça fazla olduğu Şekil-4.7'de gözükmektedir. Bilgi kazanımı için ev ve araba durumunun öneminin olmadığı gözükmektedir.

Çizelge 4.3. BK ÖS sonuçları

Nitelik	Önem Derecesi
Kategori-5	0,249048
Perşembe GSS	0,096933
Kategori-4	0,080734
Toplam GSS	0,079071
Öğleden Sonra GSS	0,072945
Kategori-3	0,053198
Çarşamba GSS	0,051289
Öğleden Önce GSS	0,048709
Akşam GSS	0,048690
Cuma GSS	0,042556
Salı GSS	0,030398
Pazartesi GSS	0,025142
Cumartesi GSS	0,020637
Kategori-2	0,020152
Meslek	0,014189
Yaşanılan Şehir	0,009371
Gece GSS	0,009086
Kategori-1	0,008900
Yaş	0,007848
Pazar GSS	0,007343
Eğitim Durumu	0,005797
Cinsiyet	0,001122
Evlilik Durumu	0,000383
Çocuğu Var Mı	0,000286
Evi Var Mı	0,000000
Arabası Var Mı	0,000000



Şekil 4.10. BK ÖS sonuç grafiği

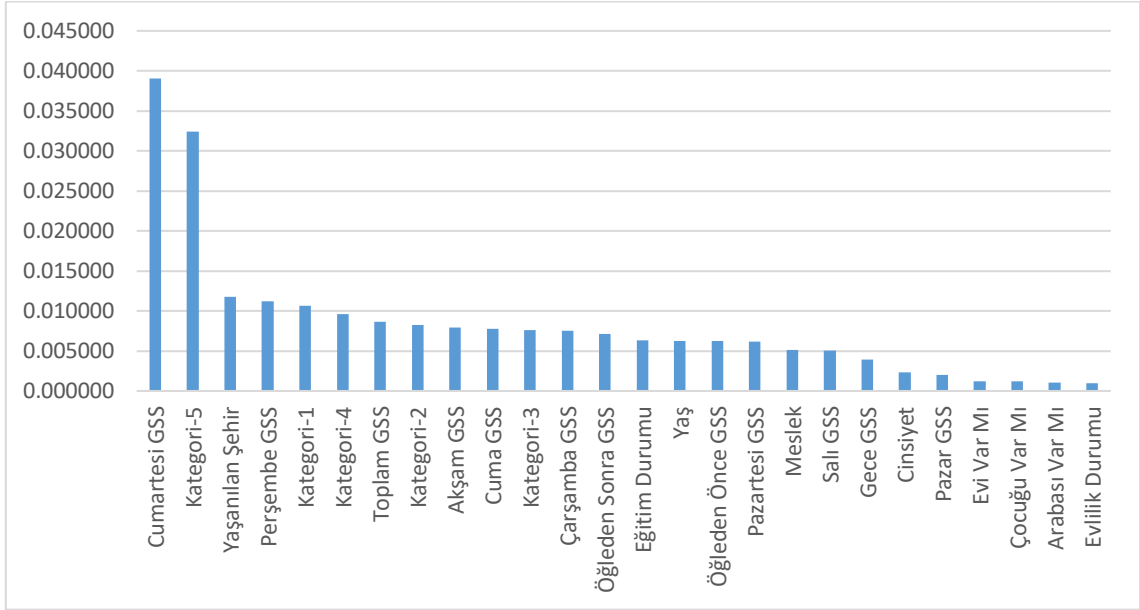
4.2.2. ReliefF özellik seçimi sonuçları

ReliefF özellik seçimi yöntemi uygulandığında elde edilen sonuçlar önem derecelerine göre Çizelge 4.4 ve Şekil 4.11’te verilmiştir. Elde edilen sonuçlar incelendiğinde önem derecesi en yüksek Cumartesi GSS niteliği ve kategori-5 niteliğinin önem dereceleri birbirine yakın çıkmıştır. ReliefF özellik seçimi için önemi olmayan nitelik bulunmamaktadır.

Çizelge 4.4. ReliefF ÖS sonuçları

Nitelik	Önem Derecesi
Cumartesi GSS	0,039042
Kategori-5	0,032400
Yaşadılan Şehir	0,011767
Perşembe GSS	0,011234
Kategori-1	0,010652
Kategori-4	0,009608
Toplam GSS	0,008640
Kategori-2	0,008245
Akşam GSS	0,007887
Cuma GSS	0,007729
Kategori-3	0,007596
Çarşamba GSS	0,007506
Öğleden Sonra GSS	0,007087
Eğitim Durumu	0,006343
Yaş	0,006273
Öğleden Önce GSS	0,006233

Pazartesi GSS	0,006187
Meslek	0,005138
Salı GSS	0,005009
Gece GSS	0,003949
Cinsiyet	0,002347
Pazar GSS	0,001963
Evi Var Mı	0,001213
Çocuğu Var Mı	0,001185
Arabası Var Mı	0,001012
Evlilik Durumu	0,000986



Şekil 4.11. Relief ÖS sonuç grafiği

4.3. Sınıflandırma

Bilgi kazanımı ve ReliefF özellik seçimi önem derecelerine göre veri seti, en önemli 5, 10, 15, 20 ve 26 (tüm nitelikler) nitelik şeklinde alt kümelere ayrılarak sınıflandırma yapılmıştır. Sınıflandırma işlemlerinde LR, YSA ve DVM algoritmaları ve 10 kat çapraz doğrulama yöntemi kullanılmıştır. Tüm veri setleri sınıflandırma algoritmalarının Çizelge 4.5’de verilen parametreleri ile sınıflandırılmıştır. Sınıflandırma işlemleri Azure ML platformu üzerinde yapılmıştır.

Çizelge 4.5. Sınıflandırma parametreleri

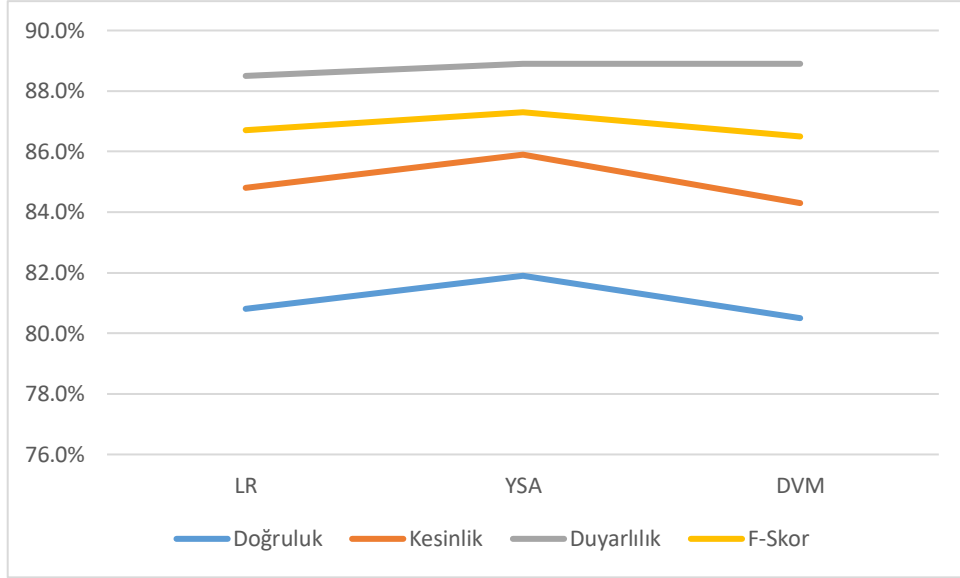
Yöntem	Parametre
LR	optimization_tolerance=1e-07
	L1_regularization_weight=1
	L2_regularization_weight=1
	memory_size_for_L-BFGS=20
	random_number_seed=0
YSA	number_of_hidden_nodes=100
	learning_rate=0,1
	number_of_learning_iterations=100
	the_initial_learning_weights_diameter=0,1
	the_momentum=0
DVM	number_of_iterations=1
	lambda=0,001
	random_number_seed=0

4.3.1. Bilgi kazanımı sınıflandırma sonuçları

BK ÖS yöntemi ile elde edilen her bir veri setinin LR, YSA ve DVM yöntemleri ile sınıflandırma sonuçları Çizelge 4.6 - Çizelge 4.10 ve Şekil 4.12 - Şekil 4.16 aralığında verilmiştir.

Çizelge 4.6. BK ÖS yöntemine göre ilk 5 niteliğin sınıflandırma sonuçları

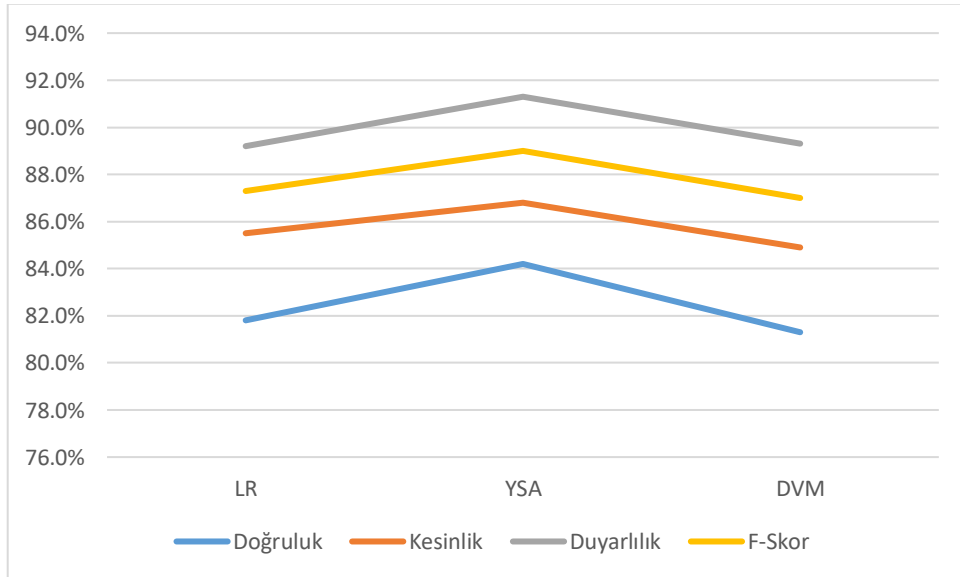
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,808	0,848	0,885	0,867	62.843	11.232	8.129	18.756	0,855
YSA	0,819	0,859	0,889	0,873	63.065	10.387	7.907	19.601	0,867
DVM	0,805	0,843	0,889	0,865	63.083	11.761	7.889	18.227	0,844



Şekil 4.12. BK ÖS yöntemine göre ilk 5 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.7. BK ÖS yöntemine göre ilk 10 niteliğin sınıflandırma sonuçları

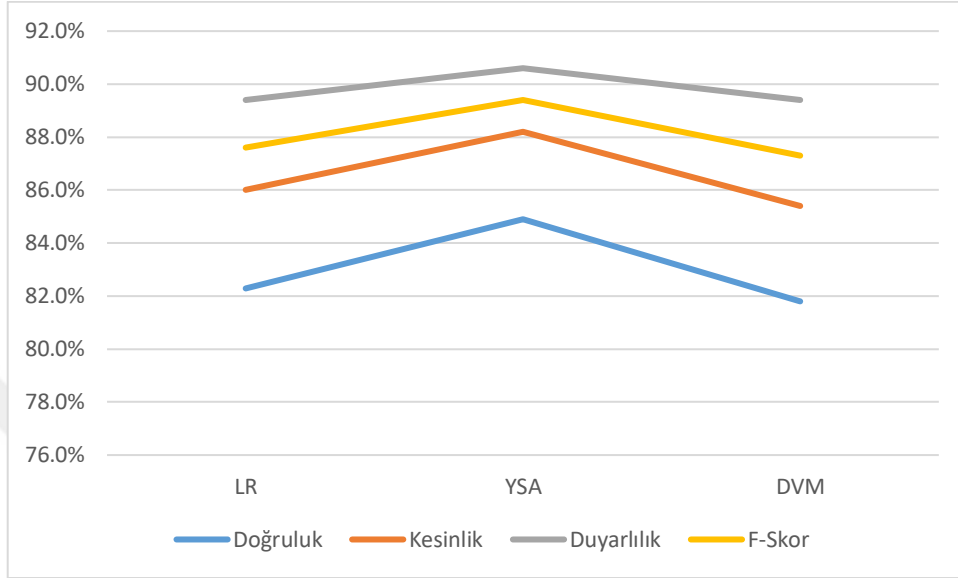
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,818	0,855	0,892	0,873	63.293	10.734	7.679	19.254	0,866
YSA	0,842	0,868	0,913	0,890	64.826	9.816	6.146	20.172	0,888
DVM	0,813	0,849	0,893	0,870	63.348	11.238	7.624	18.750	0,857



Şekil 4.13. BK ÖS yöntemine göre ilk 10 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.8. BK ÖS yöntemine göre ilk 15 niteliğin sınıflandırma sonuçları

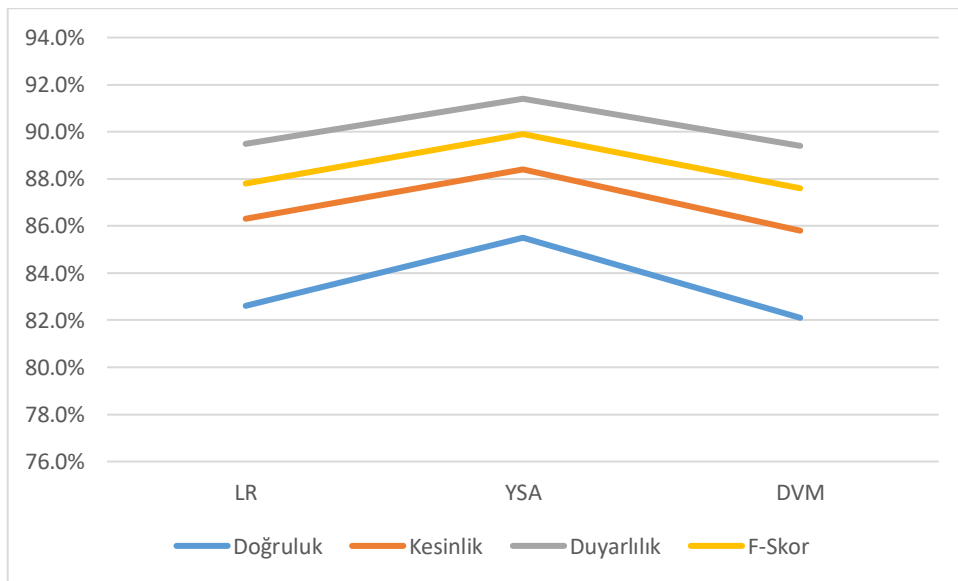
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,823	0,860	0,894	0,876	63.427	10.352	7.545	19.636	0,873
YSA	0,849	0,882	0,906	0,894	64.315	8.573	6.657	21.415	0,898
DVM	0,818	0,854	0,894	0,873	63.429	10.835	7.543	19.153	0,865



Şekil 4.14. BK ÖS yöntemine göre ilk 15 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.9. BK ÖS yöntemine göre ilk 20 niteliğin sınıflandırma sonuçları

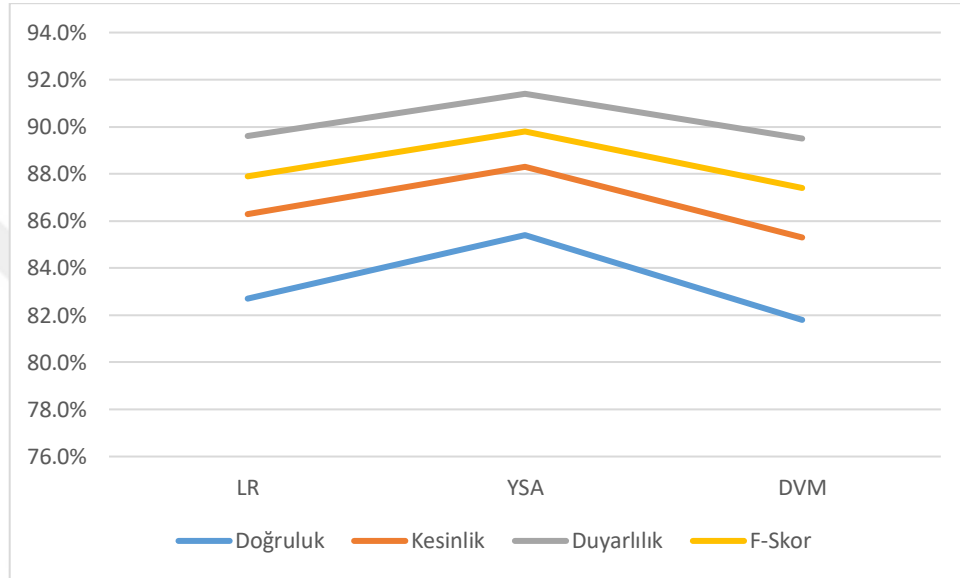
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,826	0,863	0,895	0,878	63.495	10.112	7.477	19.876	0,876
YSA	0,855	0,884	0,914	0,899	64.854	8.482	6.118	21.506	0,903
DVM	0,821	0,858	0,894	0,876	63.482	10.542	7.490	19.446	0,870



Şekil 4.15. BK ÖS yöntemine göre ilk 20 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.10. BK ÖS yöntemine göre ilk 26 niteliğin sınıflandırma sonuçları

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,827	0,863	0,896	0,879	63.596	10.114	7.376	19.874	0,878
YSA	0,854	0,883	0,914	0,898	64.837	8.629	6.135	21.359	0,902
DVM	0,818	0,853	0,895	0,874	63.521	10.927	7.451	19.061	0,868

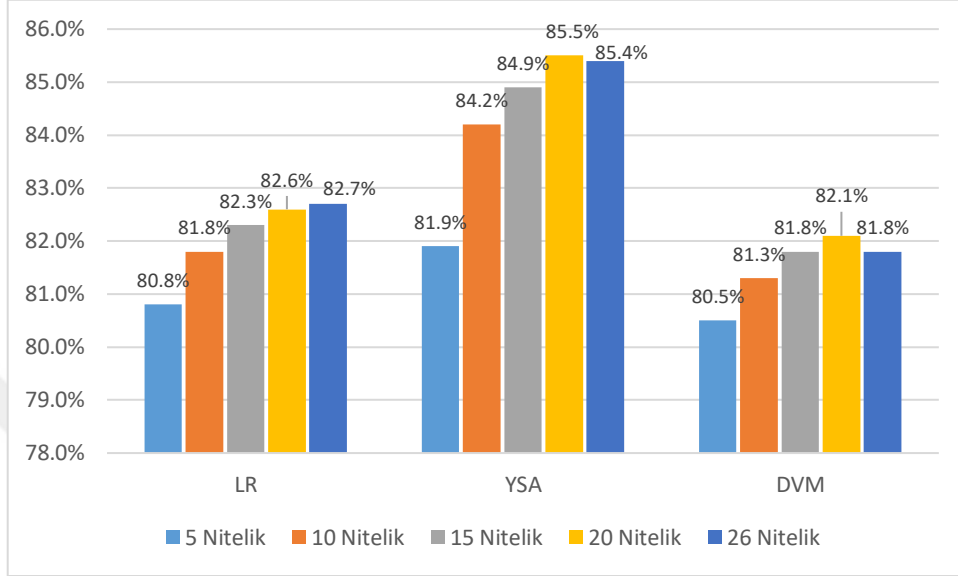


Şekil 4.16. BK ÖS yöntemine göre ilk 26 niteliğin sınıflandırma sonuç grafiği

BK ÖS yöntemi ile elde edilen her bir veri setinin LR, YSA ve DVM yöntemleri ile sınıflandırma sonucu elde edilen doğruluk oranları Çizelge 4.11 ve Şekil 4.17'de verilmiştir. Sonuçlar incelendiğinde, YSA ile elde edilen doğruluk oranının tüm veri setlerinde diğer sınıflandırma yöntemlerine göre daha iyi sonuç verdiği görülmektedir. YSA ve DVM için 20 nitelikli veri setinde en iyi doğruluk oranına ulaşılmıştır. LR için 20 ve 26 nitelikli veri setlerinde birbirine yakın doğruluk oranları çıksa da 26 nitelikte en iyi sonucu vermiştir. DVM sınıflandırma yöntemleri içinde tüm veri setlerinde en kötü sonucu vermiştir.

Çizelge 4.11. BK ÖS veri setlerinin doğruluk oranları

Yöntem	5 Nitelik	10 Nitelik	15 Nitelik	20 Nitelik	26 Nitelik
LR	0,808	0,818	0,823	0,826	0,827
YSA	0,819	0,842	0,849	0,855	0,854
DVM	0,805	0,813	0,818	0,821	0,818

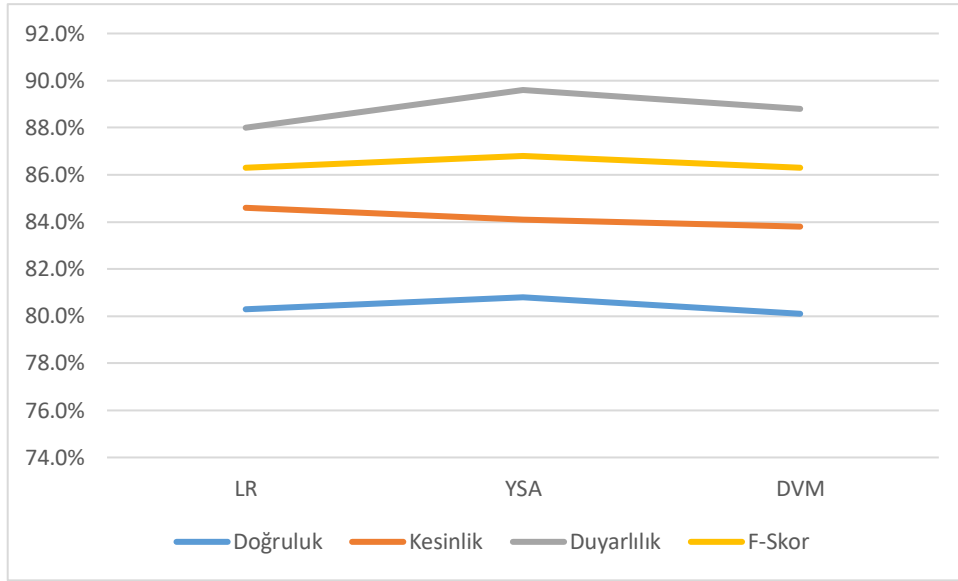
**Şekil 4.17.** BK ÖS veri setlerinin doğruluk oranları grafiği

4.3.2. ReliefF sınıflandırma sonuçları

ReliefF ÖS yöntemi ile elde edilen her bir veri setinin LR, YSA ve DVM yöntemleri ile sınıflandırılma sonuçları Çizelge 4.11 - Çizelge 4.15 ve Şekil 4.17 - Şekil 4.21 aralığında verilmiştir.

Çizelge 4.11. ReliefF ÖS yöntemine göre ilk 5 niteliğin sınıflandırma sonuçları

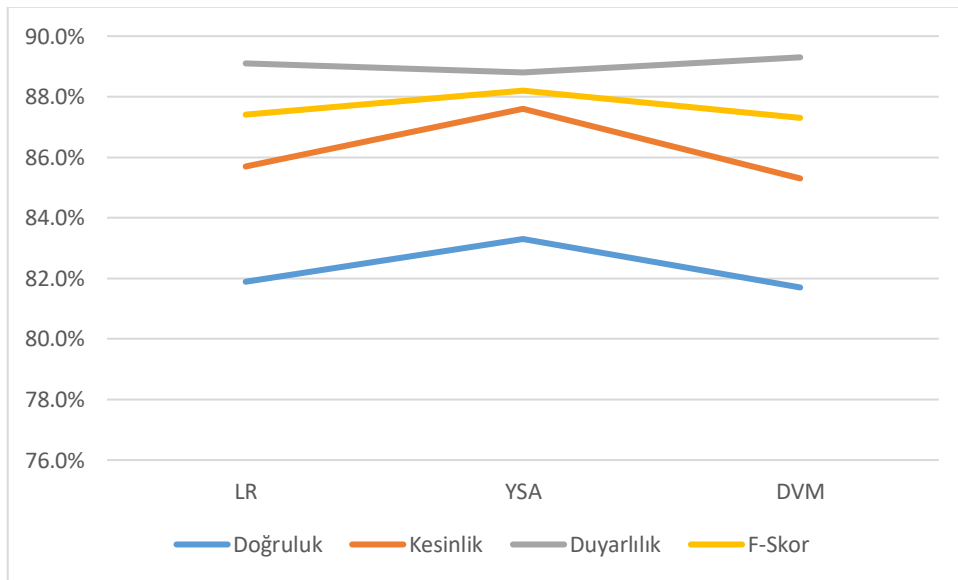
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,803	0,846	0,880	0,863	62.455	11.375	8.517	18.613	0,849
YSA	0,808	0,841	0,896	0,868	63.611	11.982	7.361	18.006	0,854
DVM	0,801	0,838	0,888	0,863	63.034	12.146	7.938	17.842	0,842



Şekil 4.17. RelifF ÖS yöntemine göre ilk 5 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.12. RelifF ÖS yöntemine göre ilk 10 niteliğin sınıflandırma sonuçları

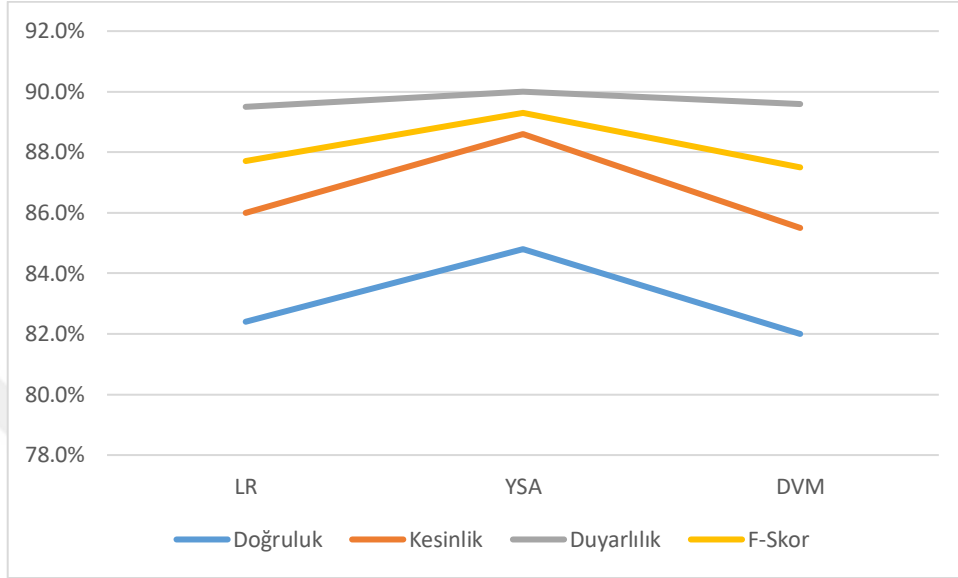
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,819	0,857	0,891	0,874	63.250	10.540	7.722	19.448	0,869
YSA	0,833	0,876	0,888	0,882	63.017	8.918	7.955	21.070	0,887
DVM	0,817	0,853	0,893	0,873	63.374	10.912	7.598	19.076	0,862



Şekil 4.18. RelifF ÖS yöntemine göre ilk 10 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.13. Relif ÖS yöntemine göre ilk 15 niteliğin sınıflandırma sonuçları

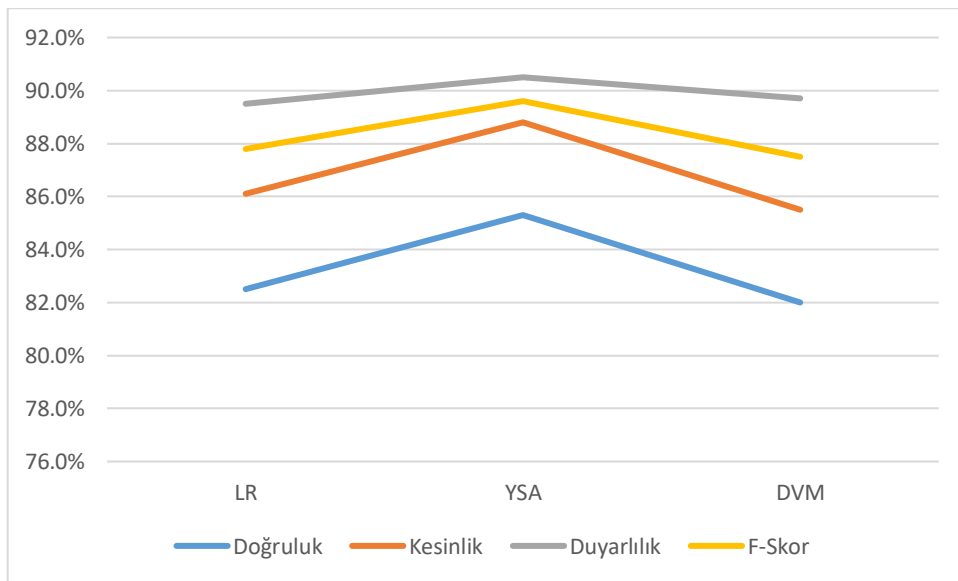
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,824	0,860	0,895	0,877	63.516	10.345	7.456	19.643	0,875
YSA	0,848	0,886	0,900	0,893	63.863	8.206	7.109	21.782	0,899
DVM	0,820	0,855	0,896	0,875	63.600	10.820	7.372	19.168	0,868



Şekil 4.19. Relif ÖS yöntemine göre ilk 15 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.14. Relif ÖS yöntemine göre ilk 20 niteliğin sınıflandırma sonuçları

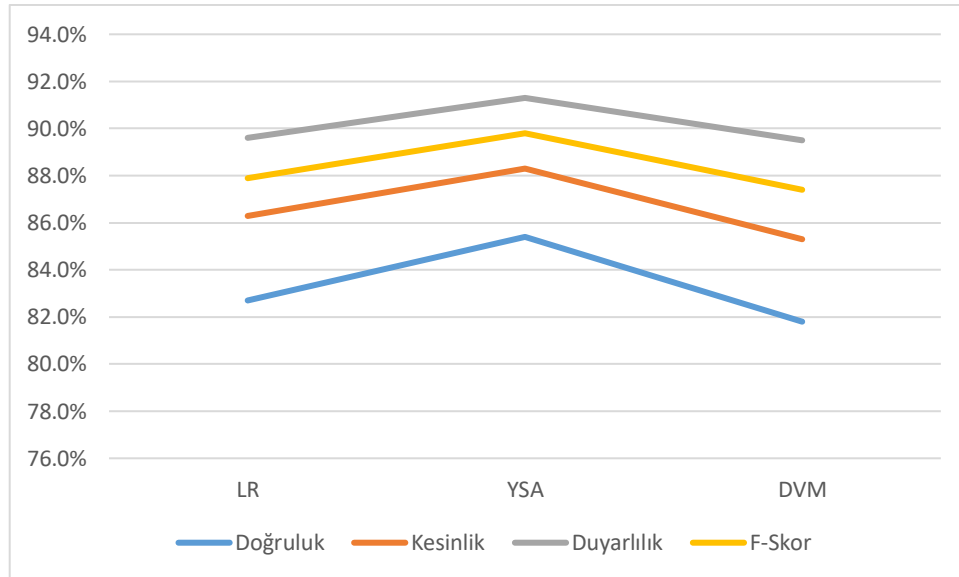
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,825	0,861	0,895	0,878	63.530	10.249	7.442	19.739	0,876
YSA	0,853	0,888	0,905	0,896	64.241	8.136	6.731	21.852	0,903
DVM	0,820	0,855	0,897	0,875	63.649	10.810	7.323	19.178	0,869



Şekil 4.20. Relif ÖS yöntemine göre ilk 20 niteliğin sınıflandırma sonuç grafiği

Çizelge 4.15. Relief ÖS yöntemine göre ilk 26 niteliğin sınıflandırma sonuçları

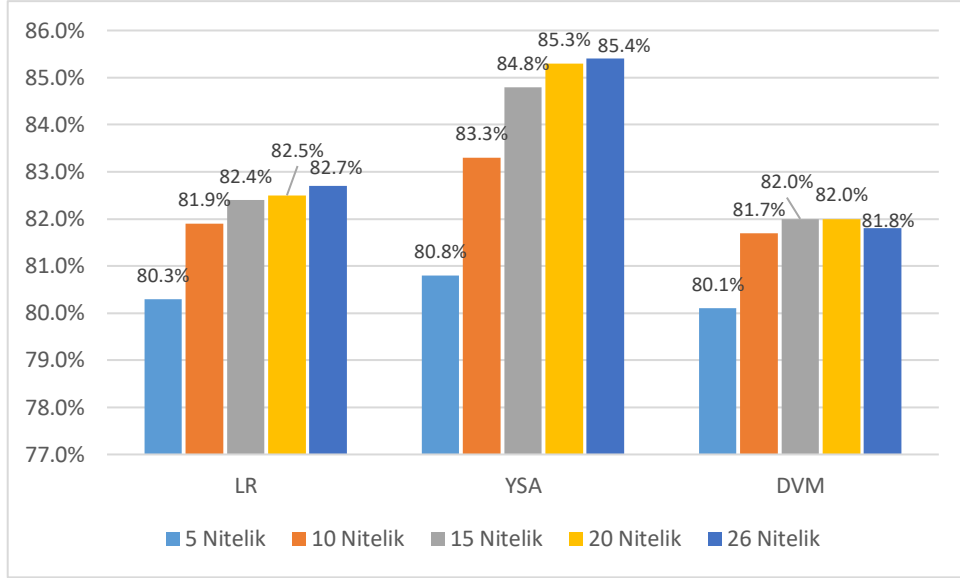
Yöntem	Doğruluk	Kesinlik	Duyarlılık	F-Skor	DP	YP	YN	DN	AUC
LR	0,827	0,863	0,896	0,879	63.599	10.112	7.373	19.876	0,878
YSA	0,854	0,883	0,913	0,898	64.807	8.595	6.165	21.393	0,902
DVM	0,818	0,853	0,895	0,874	63.521	10.927	7.451	19.061	0,868

**Şekil 4.21.** Relief ÖS yöntemine göre ilk 26 niteliğin sınıflandırma sonuç grafiği

Relief ÖS yöntemi ile elde edilen her bir veri setinin LR, YSA ve DVM yöntemleri ile sınıflandırma sonucu elde edilen doğruluk oranları Çizelge 4.16 ve Şekil 4.22’de verilmiştir. Sonuçlar incelendiğinde, YSA ile elde edilen doğruluk oranının tüm veri setlerinde diğer sınıflandırma yöntemlerine göre daha iyi sonuç verdiği görülmektedir. DVM için en iyi doğruluk oranına 15 ve 20 nitelikli veri setlerinde ulaşılmıştır. LR ve YSA için 20 ve 26 nitelikli veri setlerinde birbirine yakın doğruluk oranları olsa da 26 nitelikte en iyi sonuç vermiştir. DVM sınıflandırma yöntemleri içinde tüm veri setlerinde en kötü sonucu vermiştir.

Çizelge 4.16. Relief ÖS veri setlerinin doğruluk oranları

Yöntem	5 Nitelik	10 Nitelik	15 Nitelik	20 Nitelik	26 Nitelik
LR	0,803	0,819	0,824	0,825	0,827
YSA	0,808	0,833	0,848	0,853	0,854
DVM	0,801	0,817	0,820	0,820	0,818



Şekil 4.22. ReliefF ÖS veri setlerinin doğruluk oranları grafiği

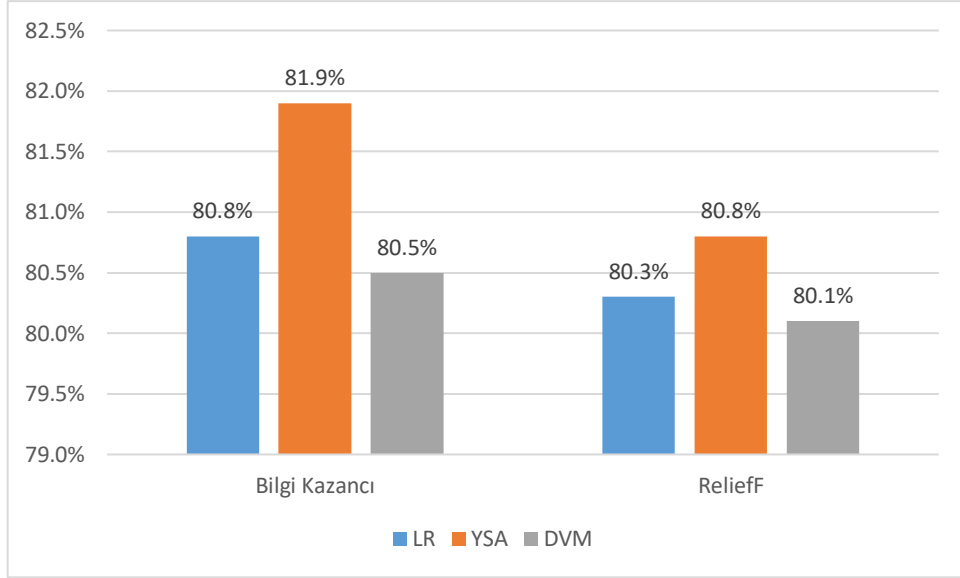
4.3.3. Sınıflandırma sonuçlarının incelenmesi

BK ve ReliefF ÖS yöntemi ile elde edilen aynı nitelikteki veri setlerinin sınıflandırılma doğruluk oranları Çizelge 4.17 - Çizelge 4.21 ve Şekil 4.23 - Şekil 4.27 aralığında verilmiştir. Tüm veri setlerinde en iyi doğruluk oranı YSA sınıflandırma yöntemi ile elde edilmiştir.

ÖS yöntemleri ile elde edilen 5 nitelikteki veri setlerinin sınıflandırma doğruluk oranları incelendiğinde, her üç sınıflandırma yönteminde BK ile elde edilen veri setlerin sınıflandırma oranlarının daha yüksek olduğu gözükmemektedir.

Çizelge 4.17. BK ve ReliefF ÖS yöntemlerine göre ilk 5 niteliğin doğruluk oranları

Yöntem	BK	ReliefF
LR	0,808	0,803
YSA	0,819	0,808
DVM	0,805	0,801

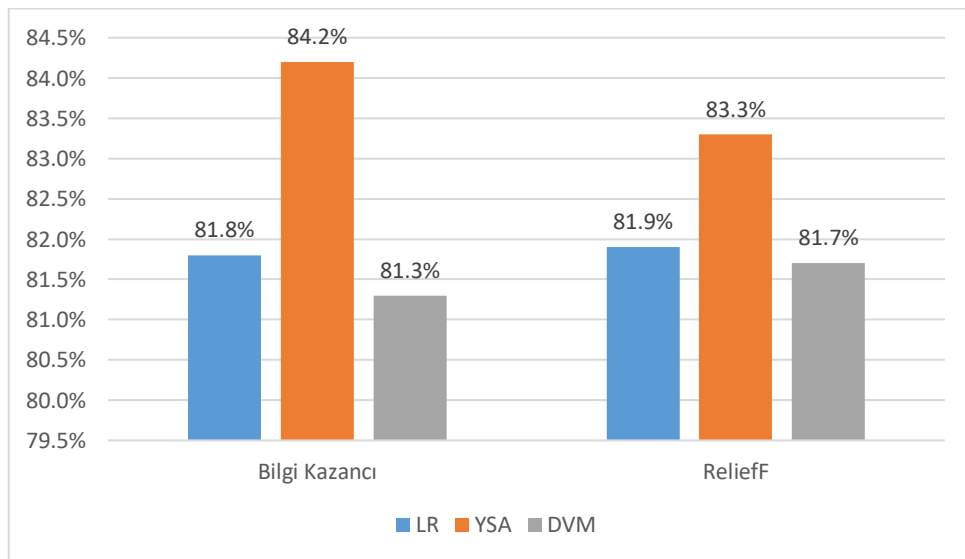


Şekil 4.23. BK ve Relief ÖS yöntemlerine göre ilk 5 niteliğin doğruluk oranları grafiği

10 nitelikteki veri setlerinin sınıflandırma doğruluk oranları incelendiğinde, LR ve DVM için Relief ile elde edilen niteliklerde doğruluk oranı daha iyi sonuç verirken YSA için BK ile elde edilen nitelikler daha iyi sonuç vermiştir.

Çizelge 4.18. BK ve Relief ÖS yöntemlerine göre ilk 10 niteliğin doğruluk oranları

Yöntem	BK	Relief
LR	0,818	0,819
YSA	0,842	0,833
DVM	0,813	0,817

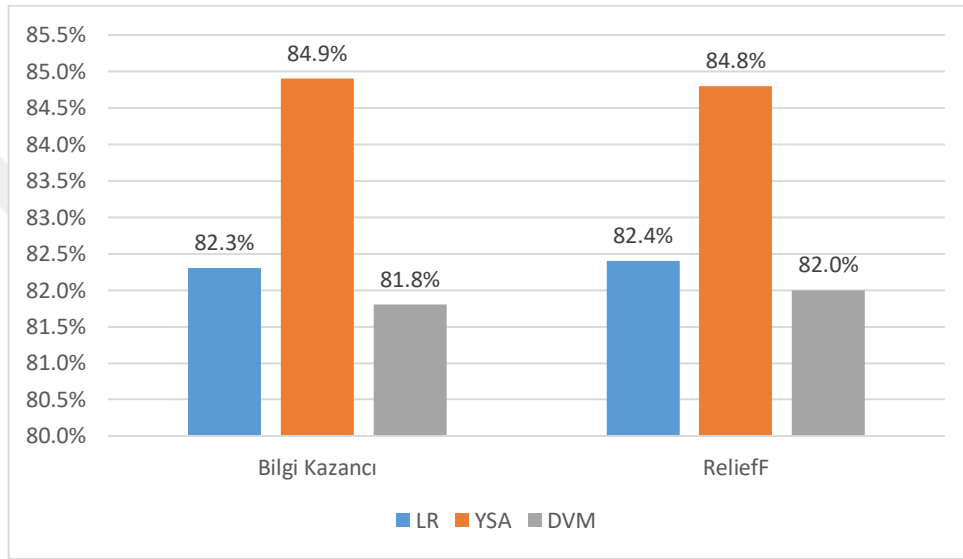


Şekil 4.24. BK ve Relief ÖS yöntemlerine göre ilk 10 niteliğin doğruluk oranları grafiği

15 nitelikteki veri setlerinin sınıflandırma doğruluk oranları incelendiğinde, LR ve YSA için BK ile elde edilen niteliklerde doğruluk oranı daha iyi sonuç verirken DVM için ReliefF ile elde edilen nitelikler daha iyi sonuç vermiştir.

Çizelge 4.19. BK ve ReliefF ÖS yöntemlerine göre ilk 15 niteliğin doğruluk oranları

Yöntem	BK	ReliefF
LR	0,823	0,824
YSA	0,849	0,848
DVM	0,818	0,820

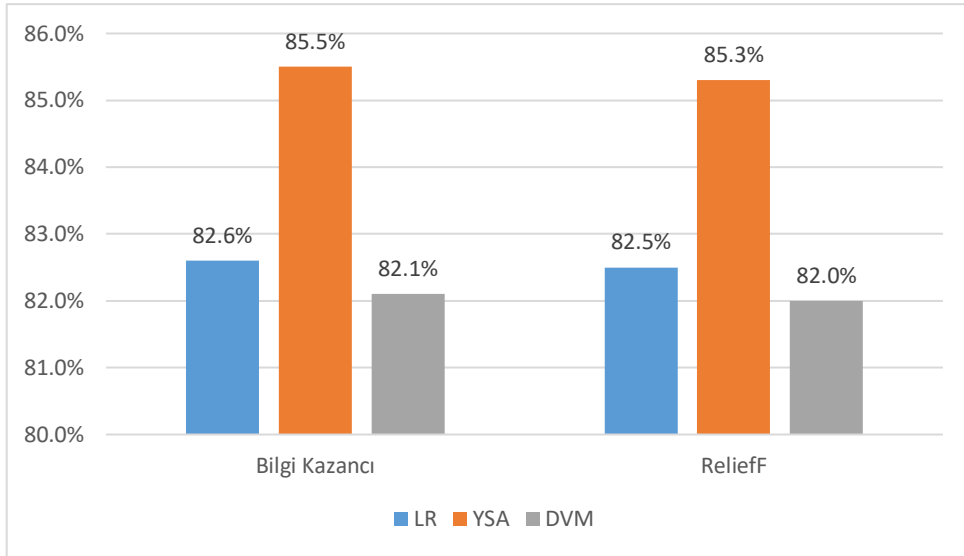


Şekil 4.25. BK ve ReliefF ÖS yöntemlerine göre ilk 15 niteliğin doğruluk oranları grafiği

20 nitelikteki veri setlerinin sınıflandırma doğruluk oranları incelendiğinde, tüm sınıflandırma yöntemleri için BK ile elde edilen nitelikler en iyi doğruluk oranını vermiştir. 20 nitelikte YSA diğer nitelik sayılarına göre en iyi sonucu vermiştir.

Çizelge 4.20. BK ve ReliefF ÖS yöntemlerine göre ilk 20 niteliğin doğruluk oranları

Yöntem	BK	ReliefF
LR	0,826	0,825
YSA	0,855	0,853
DVM	0,821	0,820

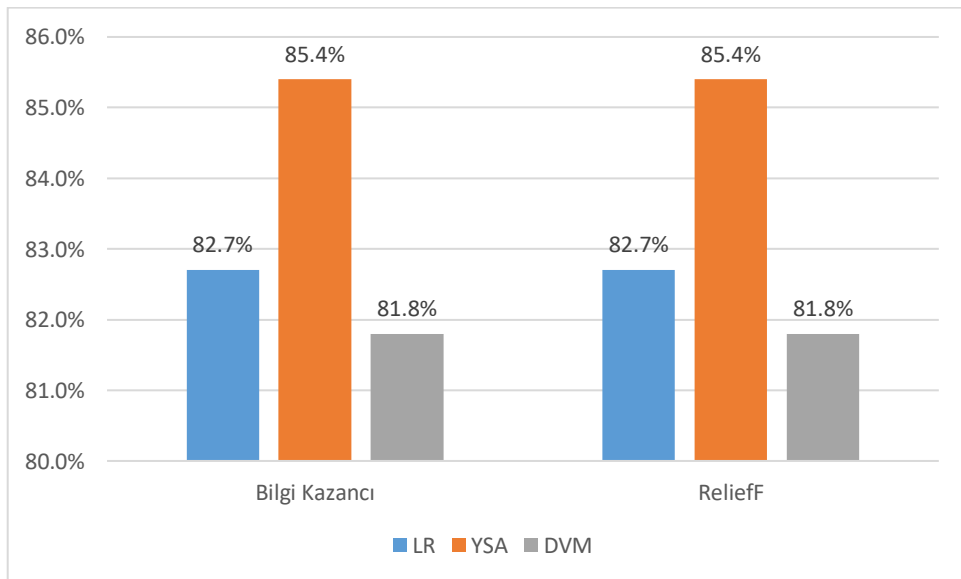


Şekil 4.26. BK ve Relief ÖS yöntemlerine göre ilk 20 niteliğin doğruluk oranları grafiği

26 nitelikteki veri setlerinin sınıflandırma işlemi incelendiğinde, veri setinde yer alan tüm niteliklerin farklı sıra ile sınıflandırıldığında aynı sonuçların elde edildiği gözlenmiştir. Buradan sınıflandırma işlemlerinde kullanılan niteliklerin sıralamalarının doğruluk üzerinde etkisinin olmadığını çıkarabiliriz.

Çizelge 4.21. BK ve Relief ÖS yöntemlerine göre ilk 26 niteliğin doğruluk oranları

Yöntem	BK	Relief
LR	0,827	0,827
YSA	0,854	0,854
DVM	0,818	0,818



Şekil 4.27. BK ve Relief ÖS yöntemlerine göre ilk 26 niteliğin doğruluk oranları grafiği

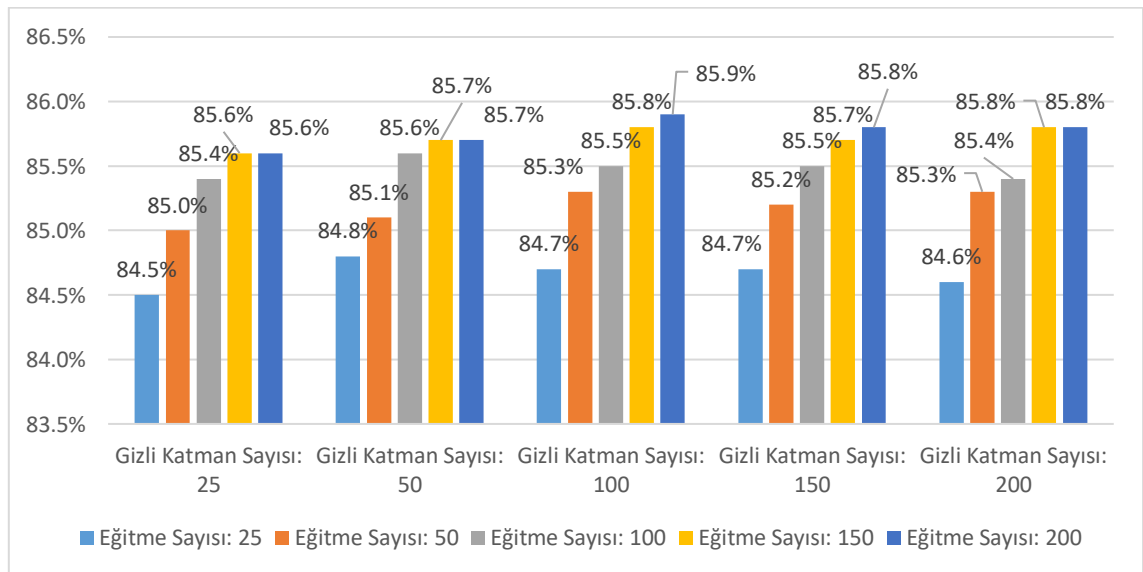
4.3.4. Yapay sinir ağı üzerine uygulama

Sınıflandırma sonuçları incelendiğinde YSA'nın en iyi doğruluk oranına BK 20 nitelik için elde ettiği gözlenmiştir. Veri seti için, YSA sınıflandırma yönteminde kullanılan bazı parametreler değiştirilerek YSA'nın doğruluğu üzerindeki etkisi incelenmiştir.

Daha önce yapılan sınıflandırma işlemlerinde YSA'nın gizli katman sayısı (number_of_hidden_nodes) ve eğitime sayısı (number_of_learning_iterations) varsayılan değeri 100 olarak sınıflandırma yapılmıştı. Gizli katman ve eğitime sayılarının YSA'nın doğruluk oranına etkisinin incelenmesi için 25, 50, 100, 150, 200 gizli katman sayılarında 25, 50, 100, 150, 200 eğitime sayılarının sınıflandırma sonuçlarına etkisi Çizelge 4.22 ve Şekil 4.28'da verilmiştir.

Çizelge 4.22. Gizli katman ve eğitime sayısının YSA'nın doğruluk oranına etkisi

		Gizli Katman Sayısı					
		25	50	100	150	200	300
Eğitime Sayısı	25	0,845	0,848	0,847	0,847	0,846	0,846
	50	0,850	0,851	0,853	0,852	0,853	0,851
	100	0,854	0,856	0,855	0,855	0,854	0,856
	150	0,856	0,857	0,858	0,857	0,858	0,859
	200	0,856	0,857	0,859	0,858	0,858	0,860
	300	0,856	0,860	0,860	0,860	0,860	0,861

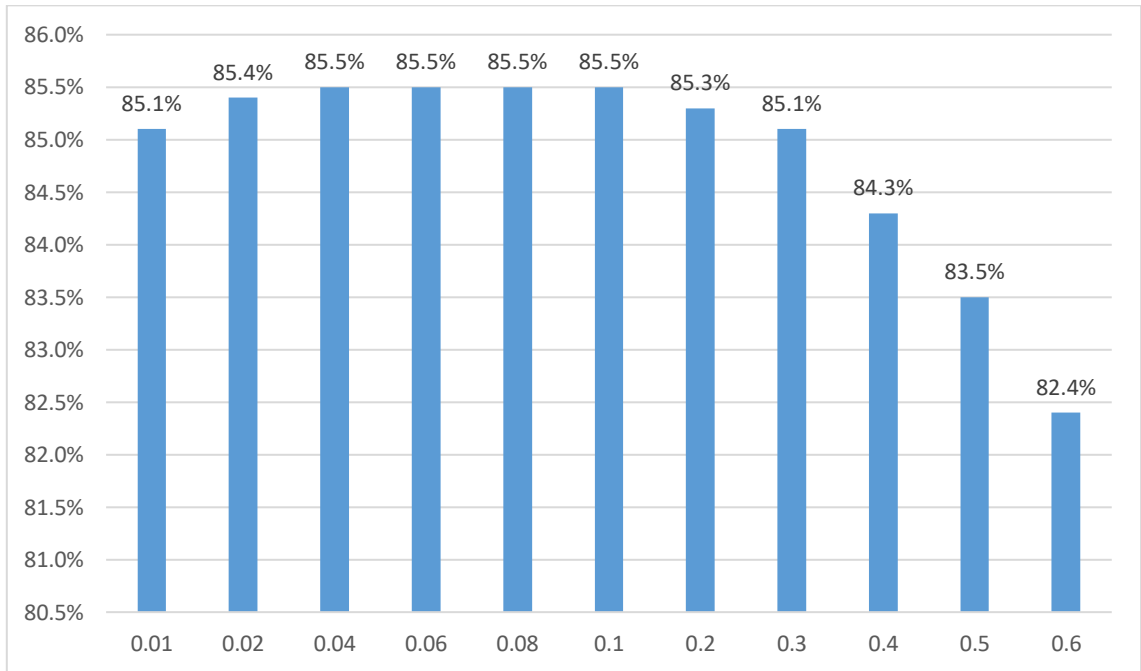


Şekil 4.28. Gizli katman ve eğitime sayısının YSA'nın doğruluk oranına etkisinin grafiği

YSA sınıflandırma işlemlerinde öğrenme hızının (learning_rate) varsayılan değeri 0,1 olarak kullanılmıştır. Öğrenme hızının doğruluk oranına etkisinin incelenmesi için farklı öğrenme hızlarında elde edilen doğruluk oranları sonuçları Çizelge 4.23 ve Şekil 4.29’da verilmiştir. Sonuçlar incelendiğinde öğrenme hızının artması doğruluk oranını azalttığı gözlenmiştir.

Çizelge 4.23. Gizli katman ve eğitme sayısının YSA’nın doğruluk oranına etkisi

Öğrenme Hızı:	0,01	0,02	0,04	0,06	0,08	0,1	0,2	0,3	0,4	0,5	0,6
Doğruluk Oranı:	0,851	0,854	0,855	0,855	0,855	0,855	0,853	0,851	0,843	0,835	0,824



Şekil 4.29. Öğrenme hızının YSA’nın doğruluk oranına etkisinin grafiği

5. SONUÇLAR VE ÖNERİLER

Bankacılık sektöründe pazarlama faaliyetleri kapsamında elde edilen büyük miktardaki verinin işlenmesi ve değerli bilgi elde edilmesi için geleneksel yöntemler yetersiz kalmaktadır. Veri madenciliği yöntemleri sayesinde büyük verilerden elde edilen yararlı bilgilerin, daha etkin kullanılması sektörde fırsatlar, müşteri memnuniyeti, yüksek karlılık, süreçlerde iyileştirmeler ve rekabet avantajı sağlanmaktadır.

Mobil pazarlama amacı ile gönderilen SMS'ler bankanın sahip olduğu ürün ve yenilikleri müşterilere bildirmesi için önemli bir iletişim ve bilgilendirme yöntemidir. Gönderilen pazarlama SMS'lerinin doğru hedef kitleye gönderilmesi verimliliğin artırılması ve müşterinin kaybedilmemesi gibi sebeplerden dolayı çok önemlidir. Ayrıca var olan müşterinin kaybedilmesinin yeni müşteri kazanmaya göre daha düşük maliyetli bir yöntem olduğu bilinmektedir.

Bu çalışma kapsamında özel bir bankanın müşterilerine göndermiş olduğu mobil pazarlama SMS'leri ve müşterilerine ait bazı demografik bilgilerinin veri madenciliği yöntemleri kullanılarak hedef kitle analizi gerçekleştirilmiştir. Hedef kitle analizinde en önemli noktalardan birisi veri setinin iyi seçilmesidir. Seçilen veri setinin gerçek verilerden oluşması, nitelik ve kayıt sayısının yeterli olması ve veri setinin ön işlemlerden geçirilmiş olması başarılı bir analiz için önemli kriterlerdir. Tez kapsamında kullanılan veri setinin özel bir bankaya ait gerçek verilerden oluşması başarılı bir analiz yapılması noktasında katkı sağlamıştır.

Uygun veri seti sağlandıktan sonra veri boyutunun azaltılması, işlem sürelerinin azaltılması ve en etkili özelliklerin tespiti için özellik seçim yöntemi olarak BK ve ReliefF özellik seçimi yöntemleri kullanılmıştır. Sınıflandırma adımında, özellik seçiminde elde edilen farklı sayılardaki veri setleri LR, YSA ve DVM sınıflandırma ve 10 kat çapraz doğrulama yöntemleri uygulanarak farklı modeller oluşturulmuştur. Elde edilen sonuçlar değerlendirildiğinde YSA diğer sınıflandırma yöntemlerine göre tüm veri setlerinde en iyi sonucu vermiştir.

Elde edilen sonuçlar değerlendirildiğinde özellik seçiminin veri boyutunun azaltılması ve etkili özelliklerin tespitinde fayda sağladığı gözlemlenmiştir. Tez çalışmasında elde edilen sonuçların bankacılık sektöründe gönderilen mobil pazarlama faaliyetlerine fayda sağlaması ve rehberlik etmesi düşünülmektedir.

KAYNAKLAR

- Akçetin, E. ve Çelik, U., 2014, İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması, *İnternet Uygulamaları ve Yönetimi Dergisi*, 5, 43-56.
- Aksu, A., 2007, Mobil Pazarlama ve Piyasa Etkinliğinin İncelenmesi, Yüksek Lisans Tezi, *Kadir Has Üniversitesi Sosyal Bilimler Enstitüsü*, İşletme Anabilim Dalı, İstanbul.
- Apampa, O., 2016, Evaluation of classification and ensemble algorithms for bank customer marketing response prediction, *Journal of international technology and information management*, 25 (4), 84-100.
- Asare-Frempong, J. ve Jayabalan, M., 2017, Predicting customer response to bank direct telemarketing campaign, *International conference on engineering technology and technopreneurship (ICE2T)*, Kuala Lumpur.
- Awad, A.I., Hassanien, A.E. ve Baba, K., 2013, Advances in Security of Information and Communication Networks, *Communications in Computer and Information Science*, Berlin, Heidelberg.
- Barutçu, S., 2008, *Mobil Pazarlama, Güncel Pazarlama Yaklaşımlarından Seçmeler*, içinde : İ. Varinli, K. Çatı (Ed.), Detay Yayıncılık, Ankara, 259-285.
- Barutçu, S. ve Öztürk Göl, M., 2009, Mobil Reklamlar ve Mobil Reklam Araçlarına Yönelik Tutumlar, *KMU İİBF Dergisi*, 11(17), 24-41.
- Bayram, N., 2017, Sosyal Bilimlerde SPSS İle Veri Analizi, *Ezgi Kitabevi*, Bursa.
- Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benitez, JM. ve Herrera, F., 2014, A Review Of Microarray Datasets And Applied Feature Selection Methods, *Information Sciences*, 282:111–135.
- Coussement, K., Lessmann, S. ve Verstraeten, G., 2016, A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry, *Decision Support Systems*, 27-36.
- Çakır, Ö., 2008, Veri Madenciliğinde Sınıflandırma Yöntemlerinin Karşılaştırılması “Bankacılık Müşteri Veri Tabanı Üzerinde Bir Uygulama”, Doktora Tezi, *Marmara Üniversitesi Sosyal Bilimler Enstitüsü*, İşletme Anabilim Dalı, İstanbul
- Çınar, A. ve Dondurmacı, G. A., 2014, Finans Sektöründe Veri Madenciliği Uygulaması, *Akademik Sosyal Araştırmalar Dergisi*, 2, 258-271.
- Davis, J. ve Goadrich, M., 2006, The relationship between PrecisionRecall and ROC curves, *23rd international Conference on Machine Learning*, Pennsylvania, USA.

- Elektronik Ticaretin Düzenlenmesi Hakkında Kanun, 2014, *Resmi Gazete (Sayı: 29166)*, <http://www.resmigazete.gov.tr/eskiler/2014/11/20141105-1.htm> [Ziyaret Tarihi: 03.02.2019].
- Elhan, A. H., 1997, Lojistik Regresyon Analizinin İncelenmesi ve Tıpta Bir Uygulaması, Yüksek Lisans Tezi, *Ankara Üniversitesi Sağlık Bilimleri Enstitüsü*, 4-29.
- Fausett, L., 1994, *Fundamentals of Neural Networks*, Prentice Hall, USA.
- Forman, G. 2003, An Extensive Empirical Study of Feature Selection Metrics for Text Classification, *Journal of Machine Learning Research*, 3, 1289–1305.
- Hacıhasanoğlu, P., 2015, Bankacılıkta Mobil Pazarlama ve Tüketici Satın Alma Kararı Üzerine Etkisi: Yozgat Örneği, Yüksek Lisans Tezi, *Bozok Üniversitesi Sosyal Bilimler Enstitüsü*, Yozgat.
- Hall, M., 1999, Correlation-based Feature Selection for Machine Learning, *The University of Waikato*, PhD Thesis, Hamilton.
- Han, J., Kamber, M., 2001, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, Massachusetts, USA, 978-0-12-381479-1.
- Hosmer, D. W. ve Lemeshow, S., 2000, *Applied Logistic Regression*, John Wiley & Sons, Inc., New York.
- İslamoğlu, E., 2015, Aralık Değerli Zaman Serilerinde Kullanılan Modelleme Teknikleri, *EÜFBED Fen Bilimleri Enstitüsü Dergisi*, 8, 178–193.
- Karaağaç, Ş. S., 2015, Churn Analysis And Churn Prediction In A Private Bank, Yüksek Lisans Tezi, *Marmara Üniversitesi Fen Bilimleri Enstitüsü*, Endüstri Mühendisliği Anabilim Dalı, İstanbul.
- Kara, Y., Boyacıoğlu, M. A. ve Baykan, Ö. K., 2011, Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange, *Expert systems with Applications*, 5311-5319.
- Kaynar, O., Tuna, M. F., Görmez, Y. ve Deveci, M. A., 2017, Makine öğrenmesi yöntemleriyle müşteri kaybı analizi, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 18 (1), 1-14.
- Kıraç, S., 2012, SMS Reklamlarına Yönelik Tüketici Tutumları Oluşturan Faktörler, Yüksek Lisans Tezi, *Bahçeşehir Üniversitesi Sosyal Bilimler Enstitüsü*, İstanbul
- Kononenko, I., 1994, Estimating attributes: Analysis and extensions of RELIEF, *Machine Learning: ECML-94*, Catania, İtalya, 171–182.
- Ladha, L. ve Deepa, T., 2011, Feature Selection Methods And Algorithms, *International Journal on Computer Science and Engineering*, 3(5), 1787-1797.

- Michael, A. ve Salter, B., 2006), *Mobile Marketing* (1st edition). New York: Butterworth-Heinemann.
- Microsoft, Azure Machine Learning Studio nedir?, <https://docs.microsoft.com/tr-tr/azure/machine-learning/studio/what-is-ml-studio> [Ziyaret Tarihi: 10.06.2019]
- Nitze, I., Schulthess, U. ve Asche, H., 2012, Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification, *Proceedings of the 4th GEOBIA, Brazil*.
- Novakavic, J., Strbac, P. ve Bulatovic, D., 2011, Toward Optimal Feature Selection Using Ranking Methods and Classification Algorithms, *Yugoslav Journal of Operations Research*, 21(1), 119-135.
- Powers, D. M., 2011, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, *Journal of Machine Learning Technologies*, 2(1), 37-63.
- Reichheld, F. F. ve Sasser, E., 1990, Zero Defections: Quality Comes to Services, *Harvard Business Review*, 68, 105-111.
- Roiger, R. J. ve Geatz, M. W., 2003, *Data Mining A Tutorial-Based Primer*, Addison Wesley, USA, 350p.
- Şeker, S. E., 2008, Entropi (Entropy, Dağınım, Dağıntı), <http://bilgisayarkavramlari.sadievrenseker.com/2008/12/17/entropi-entropy/>, [Ziyaret Tarihi: 15.03.2019]
- Şirin, E., 2017, Bir Bakışta K-Fold Cross Validation, <http://www.datascience.istanbul/2017/08/29/bir-bakista-k-fold-cross-validation/>, [Ziyaret Tarihi: 06.01.2019]
- Taşçı, A. K., 2010, Mobil pazarlama faaliyetlerinin işletmelere sağladığı katkılar ve bir uygulama örneği, Yüksek Lisans Tezi, *Maltepe Üniversitesi Sosyal Bilimler Enstitüsü*, İşletme Anabilim Dalı, İstanbul.
- Tunç, A., 2016, Finans Sektörü İçin Yapay Öğrenme Teknikleri Kullanarak Kredi Kullanabilirliğin Tespiti, Yüksek Lisans Tezi, *Selçuk Üniversitesi Fen Bilimleri Enstitüsü*, Bilgisayar Mühendisliği Anabilim Dalı, Konya.
- Usta, R., 2009, Üniversite Öğrencilerinin Mobil Reklamcılığa Karşı Tutumları, *Doğuş Üniversitesi Dergisi*, 10 (2), 294-309.
- Vapnik, V.N., 1995, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag
- Yakut, E., Elmas, B. ve Yavuz S., 2014, Yapay sinir ağları ve destek vektör makineleri yöntemleriyle borsa endeksi tahmini, *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, 19 (1), 139-157.

- Yüce, A., Gödekmerdan, L. ve Arzu, D., 2012, Tüketicilerin Mobil Pazarlama Faaliyetlerini Benimsemesi: Üniversite Öğrencileri Üzerine Bir Araştırma, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 1, 182-183.
- Watson, C., McCarthy, J. and Rowley, J., 2013, Consumer Attitudes Towards Mobile Marketing in The Smart Phone Era, *International Journal of Information Management*, 33, 840– 849.
- Witten I. H. ve Frank E., 2005, Data Mining: practical machine learning tools and techniques, 2nd ed., Morgan Kaufmann, USA.



ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Recep Dur
Uyruğu : Türkiye Cumhuriyeti
Doğum Yeri ve Tarihi : Çumra - 27.01.1990
Telefon : 544 912 8008
Faks :
e-mail : recepdur@outlook.com

EĞİTİM

Derece	Adı, İlçe, İl	Bitirme Yılı
Lise	: Çumra Anadolu Lisesi, Çumra, KONYA	2008
Üniversite	: Hacettepe Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, Çankaya, ANKARA	2014
Yüksek Lisans	: Necmettin Erbakan Üniversitesi Mühendislik ve Mimarlık Fakültesi Endüstri Mühendisliği Bölümü Meram, KONYA	Devam ediyor

İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
2014- Halen	Kuveyt Türk Katılım Bankası	Kıdemli Yazılım Mühendisi

UZMANLIK ALANI

Yazılım

YABANCI DİLLER

İngilizce

YAYINLAR