



**KREDİ TAHSİS SÜRECİNDE MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN
TAHMİN PERFORMANSLARININ KARŞILAŞTIRILMASINA YÖNELİK
BİR UYGULAMA**

HAKAN İŞBİLİR

EYLÜL 2023

ÇANKAYA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

VERİ ANALİTİĞİ ANABİLİM DALI

Yüksek Lisans

VERİ ANALİTİĞİ

KREDİ TAHSİS SÜRECİNDE MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN
TAHMİN PERFORMANSLARININ KARŞILAŞTIRILMASINA YÖNELİK
BİR UYGULAMA

HAKAN İŞBİLİR

EYLÜL 2023

ÖZET

KREDİ TAHSİS SÜRECİNDE MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN TAHMİN PERFORMANSLARININ KARŞILAŞTIRILMASINA YÖNELİK BİR UYGULAMA

İŞBİLİR, Hakan

Veri Analitiği Yüksek Lisans

Danışman: Prof. Dr. Nihat SOLAKOĞLU

Eylül 2023, 76 sayfa

Teknolojik olanakların gelişmesi, kayda değer oranda artan verinin depolanması ve söz konusu veriden stratejik analizler yapılabilmesi imkanını beraberinde getirmiştir. Veri analizi, rekabetçi piyasalarda faaliyet gösteren tüzel/gerçek kişilere karar alma aşamalarında yol gösterici değerlendirmeler yapabilme olanağı sağlamaktadır. Başta bankalar olmak üzere müşteri sayıları son yıllar içerisinde önemli ölçüde artan finansal kuruluşların, gerek mali bünyelerini sağlamlaştırabilmeleri gerekse yasal düzenlemelere uygun olarak faaliyetlerini sürdürebilmeleri amacıyla kredi tahsis kararlarında her geçen dönem daha isabetli kararlar almaları gerekmektedir. Veri madenciliği ve makine öğrenmesi yöntemleri, bankalara ve diğer finansal kuruluşlara, maruz kalabilecekleri risklere karşı ihtiyatlı hareket etme ve tedbir alma fırsatı sunmaktadır.

Çalışmamız kapsamında Lending Club isimli finans kuruluşu tarafından 2007-2020 yılları arasında tahsis edilen kredilere ilişkin bilgileri içeren veri seti üzerinde Lojistik Regresyon, Karar Ağaçları, K-En Yakın Komşu, Yapay Sinir Ağları ve LightGBM algoritmaları kullanılarak, veri setinde yer alan kredilerin tahsisi aşamasında söz konusu makine öğrenmesi yöntemlerine başvurulmuş olması durumunda ilgili kredilerden hangilerinin temerrüde düşeceğine ilişkin olarak en başarılı tahmin oranına ulaşacak algoritmanın tespit edilmesi hedeflenmiştir.

Çalışmamız sonucunda, kullanılan algoritmaların genel olarak birbirine yakın tahmin performansı gösterdiği belirlenmekle birlikte, LightGBM algoritması %81,04 doğruluk oranı elde ederek en başarılı tahmine ulaşmıştır.

Özet olarak, makine öğrenmesi uygulamaları kullanılarak elde edilecek sonuçların, finansal kuruluşlar tarafından karar alma aşamasında destekleyici birer parametre olarak dikkate alınabileceği kanaatine ulaşılmıştır.

Anahtar Kelimeler: Veri madenciliği, Makine öğrenmesi, Denetimli Öğrenme, Bankacılık, Kredi tahsisi



ABSTRACT

AN APPLICATION TO COMPARE THE PREDICTION PERFORMANCES OF MACHINE LEARNING METHODS IN THE LOAN ALLOCATION PROCESS

ISBILIR, Hakan

M.Sc. in Data Analytics

Supervisor: Prof. Dr. M. Nihat SOLAKOĞLU

September 2023, 76 pages

The development of technological possibilities has brought the opportunity to store data that has increased significantly and to make strategic analyzes from that data. Data analysis provides firms/individuals, whom operating in competitive markets, the opportunity to make strategic evaluations during the decision-making phase. Financial institutions, especially banks, whose number of customers have increased significantly in recent years, need to make more accurate decisions in loan allocation process in order to strengthen their financial structure and maintain their activities in accordance with legal regulations. Data mining and machine learning methods offer banks and other financial institutions the opportunity to act prudently and take precautions against the risks they may be exposed to.

Within the scope of our study Logistic Regression, Decision Trees, K-Nearest Neighbor, Artificial Neural Networks and LightGBM algorithms were used on the data set containing information about the loans allocated by the financial institution called Lending Club, between 2007 and 2020. It is aimed to determine the algorithm that will reach the most accurate prediction rate regarding which of the loans will default if these machine learning methods are used. As a result of this study, the algorithms performed close to each other and the LightGBM algorithm was the method that reached the most successful prediction rate with an accuracy rate of 81.04%.

In summary, the results that obtained by using machine learning applications can be used by financial institutions as a parameter in the decision-making process.

Keywords: Data mining, Machine learning, Supervised learning, Banking, Credit allocation



TEŞEKKÜR

Bu çalışma sırasında; akademik çalışmalara yönelik olarak öğrencilerine sınırsız destek sunan, yardım ve katkılarıyla beni her daim öne taşıyan, bilgilendiren ve yönlendiren değerli hocam ve tez danışmanım Prof. Dr. M. Nihat SOLAKOĞLU ile yüksek lisans çalışmalarım sırasında desteklerini her zaman hissettiğim kıymetli aileme teşekkür ederim.



İÇİNDEKİLER

İNTİHAL BULUNMADIĞINA İLİŞKİN SAYFA	iii
ÖZET.....	iv
ABSTRACT	vi
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
TABLolar LİSTESİ.....	xii
ŞEKİLLER LİSTESİ.....	xiii
KISALTMALAR LİSTESİ.....	xiv
GİRİŞ	1
BÖLÜM I	2
VERİ MADENCİLİĞİ	2
1.1 VERİ TANIMI.....	2
1.2 VERİ MADENCİLİĞİ.....	3
1.3 VERİ MADENCİLİĞİ TARİHÇESİ VE LİTERATÜR TARAMASI.....	5
1.4 VERİ MADENCİLİĞİNİN İLİŞKİLİ OLDUĞU DİĞER ALANLAR.....	6
1.4.1 Makine Öğrenmesi.....	7
1.4.2 İstatistik.....	8
1.4.3 Veri Tabanı Sistemleri ve Veri Ambarları.....	8
1.4.4 Diğer Alanlar	8
1.5 VERİ MADENCİLİĞİ SÜRECİ	9
1.6. VERİ MADENCİLİĞİNİN KULLANIM ALANLARI	11
BÖLÜM II	13
MAKİNE ÖĞRENMESİ VE MAKİNE ÖĞRENMESİ YÖNTEMLERİ.....	13
2.1 MAKİNE ÖĞRENMESİ TÜRLERİ	13
2.1.1 Denetimli Öğrenme (Supervised Learning).....	13
2.1.2 Denetimsiz Öğrenme (Unsupervised Learning)	13
2.1.3 Yarı Denetimli Öğrenme (Semi-supervised Learning).....	14
2.1.4 Aktif Öğrenme (Active Learning)	14

2.2 MAKİNE ÖĞRENMESİ METOTLARI	14
2.2.1 Sınıflandırma	15
2.2.1.1 Lojistik Regresyon	15
2.2.1.2 Karar Ağaçları.....	18
2.2.1.3 K-En Yakın Komşu (K-NN)	20
2.2.1.4 Yapay Sinir Ağları	22
2.2.1.5 LightGBM (Light Gradient Boosting Machine)	24
2.2.2 Kümeleme.....	25
2.2.3 Bağlantı Analizi Modeli (Birliktelik Kuralları).....	25
BÖLÜM III.....	26
BANKACILIKLA İLGİLİ TEMEL KAVRAMLAR	26
3.1 BANKACILIKTA KREDİ KAVRAMI	27
3.2 BANKACILIKTA TEMERRÜT VE KARŞILIK KAVRAMLARI	28
3.3 BANKACILIK SEKTÖRÜNDE KARŞILAŞILABİLECEK RİSK	
TÜRLERİ.....	32
3.4 KREDİ DEĞERLENDİRME SÜRECİNDE MEVCUT DURUM	34
BÖLÜM IV	37
BANKACILIKTA VERİ MADENCİLİĞİ İLE MAKİNE ÖĞRENMESİ	
YÖNTEMLERİNİN KULLANIM ALANLARI VE BİR UYGULAMA	37
4.1 BANKACILIKTA VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ	
YÖNTEMLERİNİN KULLANIMI.....	37
4.2 UYGULAMA	38
4.2.1 Çalışmanın Tanımlanması	39
4.2.2 Verinin Tanımlanması	41
4.2.3 Verinin Hazırlanması.....	48
4.2.4 Modelleme	48
4.2.4.1 LightGBM.....	49
4.2.4.2 Karar Ağacı	50
4.2.4.3 Lojistik Regresyon	51
4.2.4.4 K-En Yakın Komşu.....	52
4.2.4.5 Yapay Sinir Ağları	53
4.2.5 Değerlendirme	54
4.2.6 Sonuç ve Tartışma	55

KAYNAKÇA	57
ÖZGEÇMİŞ.....	61



TABLÖLAR LİSTESİ

Tablo 1: Bankacılık Sektörü Bilançosu.....	27
Tablo 2: Uygulamada Kullanılan Algoritma Sonuçları	49



ŞEKİLLER LİSTESİ

Şekil 1: Verinin ve Veri Madenciliğinin Tarihsel Gelişimi	5
Şekil 2: Veri Madenciliğinin İlişkili Olduğu Alanlar	7
Şekil 3: CRISP-DM Modeli	11
Şekil 4: Lojistik Regresyon Fonksiyonu	18
Şekil 5: Örnek Karar Ağacı	20
Şekil 6: K-NN Algoritması Örneği.....	21
Şekil 7: Yapay Sinir Ağı Modeli Şeması	23
Şekil 8: Ödenme Durumuna Göre Kredilerin Dağılımı.....	42
Şekil 9: Kredi Tutarına İlişkin Grafikler.....	43
Şekil 10: Vadelerine Göre Kredilerin Dağılımı	44
Şekil 11: Vadelerine Göre Kredilerin Temerrüt Oranı	44
Şekil 12: Faiz Oranına İlişkin Grafikler.....	45
Şekil 13: Aylık Geri Ödeme Tutarına İlişkin Grafikler	46
Şekil 14: Borçluların İlk Kredi Kullandığı Tarihlerle İlişkin Grafikler.....	47
Şekil 15: Borçlu Kredi Skorlarına İlişkin Grafikler.....	48
Şekil 16: LightGBM Karmaşıklık Matrisi	49
Şekil 17: LightGBM ROC Eğrisi	50
Şekil 18: Karar Ağacı Karmaşıklık Matrisi	50
Şekil 19: Karar Ağacı ROC Eğrisi	51
Şekil 20: Lojistik Regresyon Karmaşıklık Matrisi	51
Şekil 21: Lojistik Regresyon ROC Eğrisi.....	52
Şekil 22: K-En Yakın Komşu Karmaşıklık Matrisi.....	52
Şekil 23: K-En Yakın Komşu ROC Eğrisi	53
Şekil 24: Yapay Sinir Ağları Karmaşıklık Matrisi	53
Şekil 25: Yapay Sinir Ağları ROC Eğrisi.....	54

KISALTMALAR LİSTESİ

Kısaltmalar

BDDK	:Bankacılık Düzenleme ve Denetleme Kurumu
EKK	:En Küçük Kareler
KKB	:Kredi Kayıt Bürosu
KVKK	:Kişisel Verilerin Korunması Kanunu
USD	:Amerikan Doları
RM	:Risk Merkezi
ROC	:Receiver Operating Characteristic
TBB	:Türkiye Bankalar Birliği
TCMB	:Türkiye Cumhuriyet Merkez Bankası
TL	:Türk Lirası

GİRİŞ

1950’li yıllarda bilgi teknolojilerinin kullanılmaya başlanması bilgi toplumunun başlangıcı olarak düşünülebilir. Belirtilen döneme kadar sayım ve benzeri basit işlemler için kullanılan bilgisayarlar, veri tabanı sistemlerinin gelişimiyle birlikte veriye anlık ulaşılabilmesi, verinin işlenmesi ve stratejik karar alma aşamalarında etkin analizler yapılması amacıyla kullanılmaya başlanmıştır. Bu doğrultuda teknolojik olanaklar ekonometri, matematik ve istatistik gibi bilim dalları ile birleştirilerek veri madenciliği uygulamaları geliştirilmiştir.

Günümüzün hızla dijitalleşen dünyasında, bilim, sağlık, finans ve diğer birçok sektörde veri miktarının ve birikiminin önceki dönemlerle kıyaslanamayacak ölçüde artması nedeniyle veri üzerinde yapılan çalışmalarda izlenen geleneksel yöntemler yetersiz kalabilmektedir. Bu nedenle veri madenciliği ve makine öğrenmesi yöntemleri, veri kaynaklarından sağlıklı ve stratejik verinin temini ve analizi aşamalarında önemli birer araç haline gelmiştir.

Tez çalışmamızda bir veri madenciliği yöntemi olarak makine öğrenmesi kavramı ve makine öğrenmesi metotları üzerinde durulmuştur. Araştırmamızda öncelikle veri, veri madenciliği ve makine öğrenmesi kavramları açıklanmış, akabinde makine öğrenmesi yöntemleri detaylandırılmıştır. Takip eden bölümlerde ise bankacılıkla ilgili temel kavramlar, bankacılıkta veri madenciliği ve makine öğrenmesi yöntemlerinin uygulanması konuları izah edilmiştir. Son olarak, çalışmamızda yer verilen bilgiler ışığında örnek bir uygulama yapılmıştır. Çalışmamızda yer verilen uygulama kapsamında, sınıflandırma algoritmalarının performanslarının karşılaştırılması amacıyla “Lending Club” isimli finansman kuruluşu tarafından 2007-2020 yılları arasında kredi tahsis edilen kişilere ait veriler kullanılmıştır.

BÖLÜM I

VERİ MADENCİLİĞİ

1.1 VERİ TANIMI

Çalışmanın temel kavramı olması nedeniyle verinin literatürde yer alan bazı tanımlarının belirtilmesi gerekli görülmüştür. Veri;

- tek başına anlam taşımayan veya kullanılamayan, ayrıca enformasyona ve bilgiye temel oluşturan ilişkilendirilmeye, gruplandırılmaya, yorumlanmaya, anlamlandırılmaya ve analiz edilmeye gereksinim duyulan ham bilgidir [1].
- nesneler ve nesnelerin niteliklerinden oluşan bir kümedir. Nesne için örnek olarak kayıt, varlık ve örnek kullanılabilir [2].
- işleme sürecine dahil edilmeden, gözlem veya ölçüm yöntemleri ile ortamdan elde edilen değerlerdir [3].
- özellikle veri tabanında bulunanlar olmak üzere, enformasyon için genel bir kavramdır [4].
- yorumsuz ve içeriksiz şekiller ve/veya olgulardır. Veri; ölçüm, deney, gözlem, sayım veya araştırma ile elde edilebilmekte olup, tek başına bir anlam taşımamaktadır. Bu nedenle verinin bilgiye dönüştürülebilmesi ve anlam kazanabilmesi için bir dizi işlem sonucunda dönüştürülmesi gerekmektedir [5].

Genel bir tanım yapılması gerekirse veri; işlenmemiş durumda olan, raporlama, hesaplama ve planlamanın yanı sıra analiz yapmak amacıyla kullanılabilecek olgu olarak ifade edilebilir.

Veri madenciliğinde karşılaşılan veri türleri aşağıdaki şekilde sıralanabilir.

- Nominal (Nominal) : Nominal veri, isimlendirme amacıyla kullanılan veri türüdür. Bu nedenle veri kümesinde yer alan kişi, nesne ve kurum adları ile kategorik veriler nominal veri olarak adlandırılır. Örnek: İl adı, plaka il kodu, vb.,

- Kategorik (Categorical): Kategorik veri, nominal veri türünün özel bir şeklidir. Değerler düz metin şeklinde olmasına karşın, az sayıda ve kısıtlı bir seçenek kümesi içerisinde seçim yapılarak oluşturulur. Örnek: Cinsiyet vb.
- Sıralı (Ordinal): Verinin sırası, pozisyonu vb. bilgileri ifade etmek için kullanılan veri türüdür. Örnek: Sıra numarası, kayıt numarası, vb.
- Aralıklı (Interval): Sayısal bir değer olmakla birlikte belirli bir aralıkta olması beklenen değerlerdir. Örnek: Yaş, boy, kilo, vb.
- Oran (Ratio): Sayısal ve sürekli veri türünün özel bir halidir ve bir değer diğer bir değere oranı şeklindedir. Örnek: Vücut kitle endeksi, yoğunluk, ivme, vb.
- Sürekli (Continuous): Genellikle bir ölçümle sağlanan ve ardışık değerlerin elde edilebildiği sayısal verilerdir. Örnek: Sıcaklık, boy, vb.
- Ayırık (Discrete): Genellikle bir durumun sayısını ifade eden, sayısal olmakla birlikte, ardışık olması gerekmeyen verilerdir. Örnek: Hasta sayısı, ameliyat sayısı, vb.
- Nitel (Quality): Veriye ait kategorik ve tanımlayıcı niteliklerdir. Örnek: Göz rengi, cinsiyet, vb.
- Nicel (Quantity): Veriye ait ölçülebilir ve sayısal olarak ifade edilebilen niteliklerdir. Örnek: Boy, ağırlık, vb.

Yukarıda tanımları verilen çok sayıda veri türü olsa da, veri madenciliği yöntemleri açısından temelde sayısal ve nominal olmak üzere iki veri türü yaygın olarak kullanılmaktadır [6].

1.2 VERİ MADENCİLİĞİ

Veri madenciliği, veri yığını içerisinde önemli olan bölümün belirlenmesi veya analiz yapmak için spesifik olarak araştırılan verinin çok miktardaki veri içerisinde temin edilmesi süreci olarak tanımlanabilir. Bilgi teknolojilerinde meydana gelen gelişmelere bağlı olarak veri miktarında yaşanan kayda değer artış nedeniyle veri madenciliğinin önemi her geçen gün artmaktadır. Literatürde yer alan bazı tanımlarına yer vermek gerekirse, veri madenciliği;

- geniş ölçekli veri tabanlarında saklı bulunan eğilim ve ilişkileri tespit etme ve bunlardan ilerleyen dönemdeki gidişi belirleyecek işlem dizilerini çıkarma metodudur. Diğer bir deyişle, veri madenciliğinin amacı geçmiş

gözlem örneklerinden başlayarak bazı sonuçlar ortaya çıkarmak ve tüm ana kütleye yönelik olarak bu sonuçları mümkün olduğu kadar doğru bir şekilde genelleştirmektir [2].

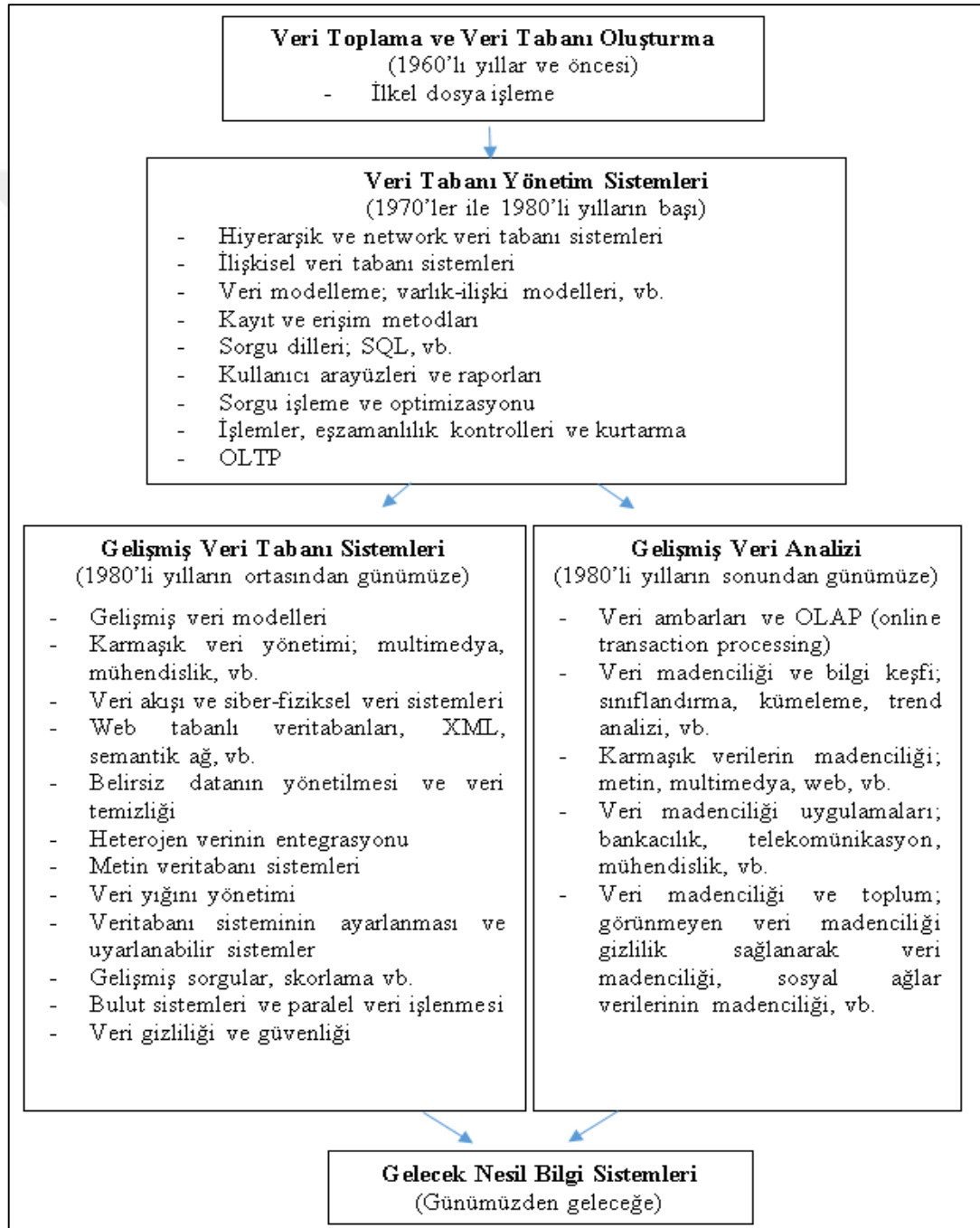
- muhtelif şekillerde ve farklı birçok kaynaktan elde edilen veriler üzerinde işlem yapılarak anlamlı bilgilere ulaşılmasıdır [3].
- büyük veri tabanlarından bilgi keşfi olarak da bilinen ve büyük veriler içerisinde bulunan ilginç, faydalı ve keşfedilmemiş örüntüler ve ilişkilerin tespit edilmesi sürecidir [6].
- büyük veri tabanlarında yer alan örtülü bilginin, tahmin edilemeyen örüntülerin ve yeni kuralların keşfedilmesi olarak ifade edilebilir [8].
- veriden bilgi keşfi sürecinin bir adımıdır; veriden örüntüler üretecek algoritmalar ve veri analizi uygulamalarından oluşur [9].
- ambarda depolanmış büyük miktardaki veriyi eleyerek anlamlı yeni korelasyonlar, örüntüler ve trendler keşfetme sürecidir [10].
- teknolojik olanakların ve farklı disiplinlerin katkısıyla, büyük hacimlerdeki veri içerisinden klasik metotlarla ortaya çıkarılması güç olan verinin anlaşılabilir nitelikte ilişkiler, örüntüler ve bağıntılar şeklinde otomatik bir biçimde ortaya çıkarılmasıdır [11].
- verilerin farklı bir açıdan analiz edilerek işe yarar bilgi şeklinde özetlenmesi sürecidir [13].
- veri tabanlarında bulunan örtülü bilgiyi ve yapıyı keşfetmek amacıyla başvurulmuş bir süreçtir. Veri madenciliği çoğu kişilerce öz bilgi keşfi ile aynı anlamda kullanılmaktadır [14].
- büyük miktarda veriden anlamlı bilgi çıkarma sanatıdır [15].
- anlam taşımayan verileri, muhtelif tekniklerle işleyerek ve analiz ederek anlaşılabilir kılan işlemlerin bütünü olarak ifade edilebilir [16].

Öte yandan, veri madenciliğinin sadece tanım olarak dikkate alınması bu alanın öneminin tam olarak anlaşılmasına sebebiyet verebilir. Yalın bir tanımda veri madenciliğine bir tahmin aracı gibi yaklaşılabılır veya veri madenciliği sıradan bir bilgisayar programı gibi algılanabilir. Şüphesiz veri madenciliği süreci bir bilgisayar yazılımı üzerinden yürütülecektir fakat veri madenciliğini sadece bir program olarak dikkate almak bu bilim alanına haksızlık olacaktır. Esasen, veri madenciliğinin kullanım alanları irdelendiğinde durumun çok daha farklı olduğu anlaşılmaktadır. Bu sebeple, ilk olarak veri madenciliğinin kullanım alanlarının ve amaçlarının

öğrenilmesi, akabinde bu amaçlar doğrultusunda geliştirilmiş algoritmaları ve yazılımları kullanmak yerinde olacaktır [7].

1.3 VERİ MADENCİLİĞİ TARİHÇESİ VE LİTERATÜR TARAMASI

Veri madenciliği, bilgi teknolojisi sürecinin bir neticesi olarak dikkate alınabilir. Verinin ve veri madenciliğinin tarihsel gelişimindeki aşamalarını aşağıdaki şekilde ele alabiliriz.



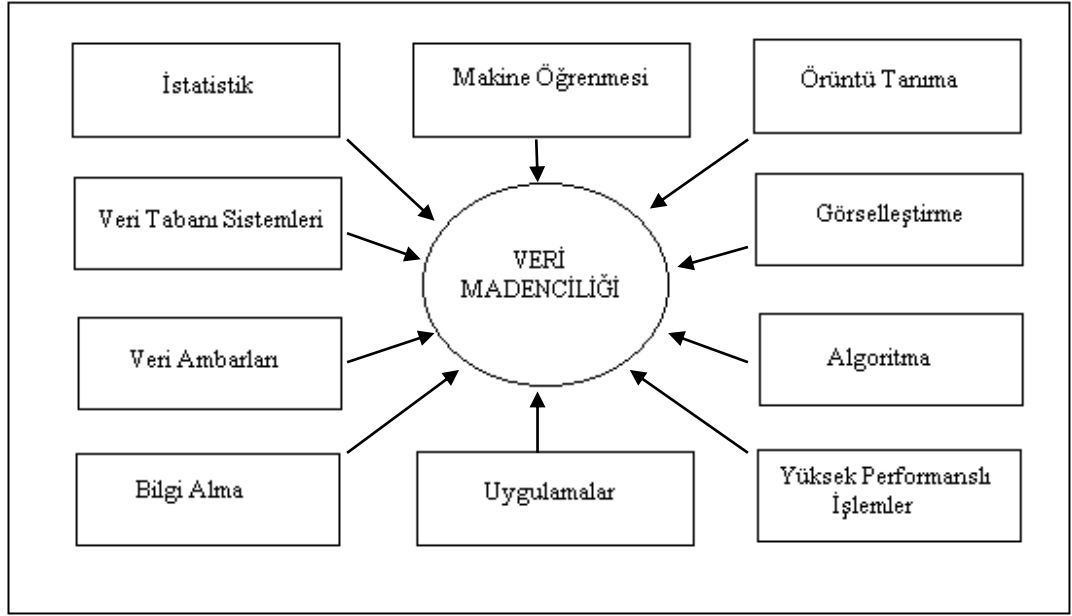
Şekil 1: Verinin ve Veri Madenciliğinin Tarihsel Gelişimi [17]

Bilgisayarlar 1950’li yıllarda sayım amacıyla kullanılmaya başlanmıştır. 1960’lara gelindiğinde ise veri tabanı ve verilerin depolanması kavramları bilişim dünyasına dahil olmuştur. 1960’ların sonuna doğru ise basit öğrenmeli bilgisayarlar geliştirilmiş olup, 1970’lerde “İlişkisel Veri Tabanı Yönetim Sistemleri” uygulamaları kullanılmaya başlanmıştır. Değinilen alanda görev yapan uzmanlar basit kurallar üzerine inşa edilen sistemler kurmuş ve bu sayede makine öğreniminin temelleri atılmıştır. 1980’li yıllarda veri tabanı yönetim sistemlerinin kullanım alanları genişlemiş ve birçok sektörde kullanılmaya başlanmıştır. Söz konusu dönemde işletmeler, müşterilerine, rakiplerine ve ürünlerine ilişkin bilgilerden meydana gelen veri tabanları oluşturmuşlardır. Değinilen veri tabanları kayda değer ölçüde veri içermekte olup, bu verilere SQL veri tabanı sorgulama dili ya da benzer diller aracılığıyla erişilebilir. 1990’lı yıllara gelindiğinde, barındırdığı veri miktarı katlanarak artan veri tabanlarından işe yarayacak bilginin bulunmasına ilişkin yöntemler araştırılmaya başlanmış ve 1992 yılında veri madenciliğinde kullanılabilecek ilk yazılım geliştirilmiştir. 2000’li yıllarda veri madenciliği gelişmeye devam etmiş ve birçok sektörde kullanılır hale gelmiştir. Elde edilen neticelerin faydaları anlaşıldıkça veri madenciliğine yönelik ilgi artmıştır [12].

Yukarıda yer verilen aşamalardan da görüleceği üzere, veri tabanı ve bilgi teknolojileri 1950’li yıllardan itibaren kayda değer ölçüde gelişim sağlamıştır. Ayrıca, veri tabanı yönetimi sistemlerindeki olumlu gelişmeler veri madenciliğine önemli katkıda bulunmuştur.

1.4 VERİ MADENCİLİĞİNİN İLİŞKİLİ OLDUĞU DİĞER ALANLAR

Uygulama ağırlıklı bir alan olan veri madenciliği; istatistik, makine öğrenimi, örüntü tanıma, veri tabanı, veri tabanı sistemleri, bilgi alma, görselleştirme, algoritmalar, yüksek performanslı bilgi işlem ve birçok uygulama tekniğini bünyesinde barındırmaktadır (Şekil-2). Disiplinler arası araştırma ve geliştirme niteliği veri madenciliğinin başarısına ve geniş uygulama alanına kayda değer oranda katkı sağlamaktadır [17].



Şekil 2: Veri Madenciliğinin İlişkili Olduğu Alanlar [17]

1.4.1 Makine Öğrenmesi

Makine öğrenmesi, hesaplama yöntemleri aracılığıyla deneyimlerden öğrenerek sistem performansını artıran bir tekniktir. Bilgisayar sistemlerinde deneyim veri biçimindedir ve makine öğreniminin temel amacı, verilerden modeller meydana getiren öğrenme algoritmaları geliştirmektir. Bu sayede öğrenme algoritmalarını deneyim verileriyle besleyerek, yeni gözlemler üzerinde tahmin yapabilen modeller elde edilebilmektedir [18].

Makine öğrenimi, bilgisayarların net biçimde programlanmadan öğrenme kabiliyeti kazanmalarını sağlayan bir alan olarak tanımlanır. Makine öğrenimi yöntemlerine, makineler verileri daha etkin biçimde kullanabilmelerini öğretmek amacıyla başvurulur. İnceleme aşaması sonrasında verilerin yorumlanmasında güçlüklerle karşılaşıldığı durumlarda makine öğrenimi uygulanabilmektedir. Mevcut veri setlerinin sayısının artmasına bağlı olarak makine öğrenimine yönelik talep artmaktadır. Makine öğreniminin amacı verilerden öğrenmektir. Açıkça programlanmaksızın makinelerin kendi kendine öğrenmesinin nasıl sağlanabileceği konusunda birçok çalışma yapılmıştır. Ayrıca, birçok matematikçi ve programcı, büyük veri setlerine sahip olan bu sorunun çözümünü bulmak için çeşitli yaklaşımlar uygulamaktadır [19].

1.4.2 İstatistik

Veri madenciliğinin, temin edilen verinin analizi, yorumlanması ve açıklanması ile ilgilenen istatistik bilimi ile yoğun ve doğal bir bağlantısı vardır.

İstatiksel model, hedef sınıfındaki nesnelerin davranışını rastgele değişkenler ve bunlarla ilişkili olasılık dağılımları yönünden açıklayan ifade eden matematiksel fonksiyonlardır. Verileri ve veri sınıflarını modellemek amacıyla istatistiksel modellerden sıklıkla faydalanılır. Örneğin, istatistik kullanılarak eksik veri değerleri modellenenbilir ve akabinde veri havuzundaki kalıpları incelerken veri madenciliği aşamaları gürültülü veya eksik değerleri tanımlamak ve işlemek için söz konusu modeli kullanabilir [17].

1.4.3 Veri Tabanı Sistemleri ve Veri Ambarları

Veri tabanlarının oluşturulması, bakımı ve kullanımı alanlarına odaklanan veri tabanı sistemleri genellikle büyük veri kümelerini işlemedeki yüksek ölçeklenebilirlikleri ile bilinir.

Veri madenciliği işlemlerinin kayda değer bölümünün büyük veri kümelerini ve hatta gerçek zamanlı, hızlı akış verilerini işleyecek şekilde tasarlanması gerekir. Bu sebeple veri madenciliği, büyük veri kümelerinde yüksek verimliliğe ve ölçeklenebilirliğe erişebilmek için veri tabanı teknolojilerinden faydalanılabilir.

Güncel veri tabanı sistemleri, veri ambarı ve veri madenciliği imkanları sayesinde veri tabanı üzerinde sistematik veri analizi yetenekleri oluşturmuştur. Bir veri ambarı, birden fazla kaynaktan ve çeşitli zaman dilimlerinden gelen verileri entegre eder [17].

1.4.4 Diğer Alanlar

Bilgi alma, belgeleri ve belgelerdeki bilgileri arama bilimidir. Belgeler metin veya medya olarak internet ortamında yer alabilir.

Örüntü tanıma, devamlı olarak yinelenen şekiller olan örüntüleri tanıma amacıyla kullanılan bir teknoloji türüdür. Örneğin, insan yüzü tanınması, parmak izi, vb. yöntemler gösterilebilir.

Görselleştirme, işletmelerin verilerden grafik form, çizim ve diğer görsel sunum türleriyle anlam çıkarmalarını sağlayan bir tekniktir. Görselleştirme sayesinde verinin daha kolay anlaşılması ve analiz edilmesi olanağı elde edilir.

Algoritma, belirlenen problemi çözebilmek veya bir amaca erişebilmek için tasarlanan süreç şeklinde ifade edilebilir.

Bir bütün olarak değerlendirildiğinde veri madenciliği, yukarıda değinilen alanlar ile etkileşimli biçimde yürütülerek hedeflenen sonuçlara ulaşılmasına katkı sağlamaktadır.

1.5 VERİ MADENCİLİĞİ SÜRECİ

1996 yılına kadar veri madenciliği sürecine yönelik genel kabul görmüş bir model bulunmamaktaydı. Bu nedenle, veri madenciliği süreci modelinin standardize edilebilmesi ve kuruluşlara kendi veri madenciliği projelerini yapabilmelerini sağlayabilmeleri amacıyla bir veri madenciliği süreci modeli ihtiyacı bulunmaktaydı. Tescil zorunluluğu bulunmayan, yazılı hale getirilmiş ve ücretsiz bir modelin, kullanıcıların veri madenciliği sürecinden elde ettiği faydayı artırmaya, sektördeki en iyi uygulamalara yönlendirmesine ve pazarın olgunlaşmasına olanak sağlama imkanı bulunmaktaydı. 1996 yılının sonunda veri madenciliği sektöründe yer alan 4 lider kuruluş (Daimler-Benz, Integral Solutions Ltd., NCR ve OHRA), CRISP-DM (Cross-Industry Standard Process for Data Mining) modelini geliştirdi.

Basit bir şekilde tanımlamak gerekirse CRISP-DM, veri madenciliği alanındaki uzmanlığına bakılmaksızın herkes tarafından kullanılabilen, veri madenciliği projesinin bir bütün olarak yönetimini sağlamaya olanak tanıyan bir veri madenciliği metodolojisi ve süreç modeli olarak ifade edilebilir [20].

CRISP-DM modeline göre veri madenciliği sürecinde takip edilen aşamalar genellikle aşağıda belirtildiği şekilde sıralanabilir:

i. Çalışmanın Tanımlanması

Veri madenciliği sürecinin muhtemelen en kritik aşaması olan çalışmanın tanımlanması (Business Understanding), işletme açısından çalışmanın hedeflerinin anlaşılması, bu hedefin bir veri madenciliği problemine dönüştürülmesi ve belirlenen hedefe ulaşılabilmesi için bir taslak plan hazırlanması olarak ifade edilebilir.

Bu aşamada özetle, bir problemin neden kaynaklandığı, problemin çözümüne ilişkin beklentiler ile problemin ilişkili olduğu veri kaynakları belirlenir ve problemin çözülmesi sonrasında elde edilecek çıktılar tanımlanır. Örneğin, kuraklık yaşanması durumunda ülkenin tarım üretimindeki azalmanın tarım ürünü satan bir firmayı hangi ölçüde etkileyeceği, vb.

ii. Verinin Tanımlanması

Verinin toplanması ile başlayan bu süreç, verinin anlaşılması ve veri kalitesinin değerlendirilmesi aşamalarını kapsar. Bu aşamada veri tipi belirlenerek, verinin gürültülü veya kirli olması, eksik veri barındırması ve benzeri problemler tespit edilir.

iii. Verinin Hazırlanması

Verilerin hazırlanması süreci ham veriden, nihai veri setinin veya modelleme aşamasında kullanılacak araçların oluşturulması aşamasını ifade etmektedir. Verilerin hazırlanması verinin seçilmesi, temizlenmesi, inşa edilmesi, birleştirilmesi ve dönüştürmesi aşamalarını kapsamaktadır.

iv. Modelleme

Tanımlanan problem için en uygun modelin bulunabilmesi amacıyla birçok model kurulur ve optimal değerlere ulaşılan kadar parametreler üzerinde değişiklik yapılır.

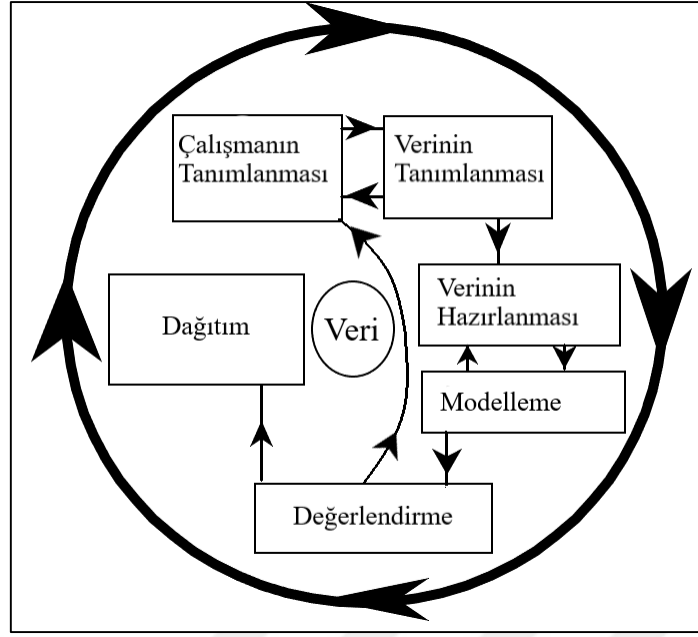
v. Değerlendirme

Bu süreç, modelin nihai halinin verilmesi öncesinde, modelin değerlendirilmesinin ve inşa aşamasının işletme amacına uygun olup olmadığının gözden geçirilmesi aşamasıdır.

vi. Dağıtım

Dağıtım aşamasında, süreç boyunca elde edilen verinin düzenlenmesi ve talep eden kişinin kullanabileceği biçimde sunulması gerekmektedir.

CRISP-DM modeline ilişkin aşamalar aşağıdaki şemada ayrıca belirtilmiştir.



Şekil 3: CRISP-DM Modeli [20]

Veri madenciliği sürecindeki aşamalarda yaşanacak olumsuzlukların, takip eden adımlarda zaman ve kaynak kaybına yol açma ihtimali yüksektir. Bu nedenle, her aşamada takip edilmesi gereken sürece uyulması sürecin başarılı sonuçlanması açısından elzemdir.

1.6. VERİ MADENCİLİĞİNİN KULLANIM ALANLARI

Günümüzde veri madenciliği işletmeler tarafından ağırlıklı olarak pazarlama, iletişim ve benzeri alanlarda müşteri odaklılığı artırmak amacıyla kullanılmaktadır. Veri madenciliği işletmelere fiyat, üretim seviyelerinin belirlenmesi, personel yetkinlikleri ve benzeri kurum içi faktörleri tespit edebilmelerine imkan sağlamaktadır. İlaveten, ekonomik veriler, rekabet ve faaliyette bulunulan piyasanın yapısı gibi dış unsurların belirlenmesine olanak tanımaktadır. Bu sayede, işletmelerin satış miktarları, müşterilerinin memnuniyeti ve şirketin elde ettiği sonuçlar üzerindeki olumlu veya olumsuz etkiler tespit edilebilmektedir. Netice itibarıyla veri madenciliği süreci, ham bilgiyi edinme ve veriler içerisindeki detayları fark etme imkanı sağlamaktadır. Değinilen avantajları sayesinde, veri madenciliği günümüzde hemen her alanda kullanım alanı bulabilmektedir. Veri madenciliğine sıklıkla başvuru alan sektörler;

- sağlık,
- biyoloji ve tıp,
- telekomünikasyon,

- bankacılık ve finans,
- sigorta,
- pazarlama,
- eğitim,
- astronomi

olarak özetlenebilir [13].

Her geçen gün daha yaygın kullanılan veri madenciliği, uygulanabilirliğinin kolay olması ve etkili sonuçlar vermesi sayesinde işletmeler tarafından sıklıkla başvurulan yöntemlerden bir tanesi haline gelmiştir [12].



BÖLÜM II

MAKİNE ÖĞRENMESİ VE MAKİNE ÖĞRENMESİ YÖNTEMLERİ

Bir veri madenciliği yöntemi olarak makine öğrenmesi, kullanılan veriler üzerinden öğrenebilen veya performansı geliştiren sistemler inşa etmeye yoğunlaşan bir yapay zeka dalıdır. Yapay zeka, insan zekasını taklit eden sistemler veya makineler bütünü anlamı taşıyan bir kavramdır. Makine öğrenimi ve yapay zeka genellikle bir arada değerlendirilir. Bu iki kavram zaman zaman birbirinin yerine kullanılsa da aynı anlamı taşımamaktadırlar. Makine öğrenmesi çözümlerinin tamamı yapay zeka olmasına karşın, tüm yapay zeka çözümleri makine öğrenimi değildir. Geline nokta makine öğrenmesi birçok yerde uygulanmaktadır. Bankalarla irtibat kurulduğunda, internet üzerinden alışveriş yapıldığında veya sosyal medya kullanıldığında makine öğrenimi algoritmaları devreye girmektedir [21].

2.1 MAKİNE ÖĞRENMESİ TÜRLERİ

Hızlı gelişen bir disiplin olan makine öğrenmesi türleri aşağıdaki başlıklar altında sınıflandırılabilir.

2.1.1 Denetimli Öğrenme (Supervised Learning)

Denetimli öğrenme sınıflandırma ile eş anlamlıdır. Öğrenme, eğitim veri seti içerisinde yer alan etiketli örneklerden sağlanır. Örnek vermek gerekirse, önceki dönemde kredi kullandırılan müşterilerin verileri üzerinden öğrenme aşaması tamamlandıktan sonra, yeni başvuruda bulunan kişilerin kredilerini ödeyememe ihtimalinin hesaplanması denetimli öğrenme örneği olarak gösterilebilir.

2.1.2 Denetimsiz Öğrenme (Unsupervised Learning)

Denetimsiz öğrenme esasen kümeleme ile aynı anlamı taşımaktadır. Girdi örnekleri sınıf etiketi içermez ve bu sebeple öğrenme süreci denetimsizdir. Genel olarak verilerde yer alan sınıfları ortaya çıkarmak amacıyla kümeleme kullanılabilir. Bir örnekle izah etmek gerekirse denetimsiz bir öğrenme metodu girdi olarak el yazısı

rakamlarından oluşan bir takım görüntü kullanabilir. 5 veri kümesinin bulunduğu senaryoda kümeler, sırasıyla 0 ile 4 arasındaki 5 farklı basamağa denk gelebilir. Diğer taraftan, eğitim verilerinin etiketlenmemesi nedeniyle, öğrenilen model bulunan kümelerin anlamsal karşılığını ifade edemeyecektir.

2.1.3 Yarı Denetimli Öğrenme (Semi-supervised Learning)

Öğrenme aşamasında etiketli ve etiketsiz örnekleri kullanan makine öğrenimi tekniklerine yarı denetimli öğrenme adı verilir. Bu yaklaşımda sınıf modellerini öğrenmek için etiketli örneklerden faydalanılır ve sınıflar arasındaki sınırları belirgin hale getirmek amacıyla etiketsiz örneklerle başvurulur.

2.1.4 Aktif Öğrenme (Active Learning)

Kullanıcıların öğrenme sürecinde aktif rol oynadığı makine öğrenmesi yöntemlerine aktif öğrenme metotları adı verilir. Aktif öğrenme yaklaşımı kullanıcılardan, etiketlenmemiş örneklerden veya öğrenme programından analizi tamamlanmış bir örneği etiketlemesini talep edebilir. Anılan uygulamadaki amaç, etiketlemelerinin istenebileceği örnek sayısına ilişkin kısıtlama bulunduğu durumlarda, kullanıcılardan aktif biçimde bilgi alarak model kalitesini optimize etmektir.

2.2 MAKİNE ÖĞRENMESİ METOTLARI

Makine öğrenmesinde kullanılan yöntemler ve algoritmalar gün geçtikçe artmaktadır. Makine öğrenmesi modelleri esasen;

- Sınıflandırma (Classification),
- Kümeleme (Clustering),
- Bağlantı Analizi Modeli (Birliktelik Kuralları, Association Rules)

olmak üzere 3 grupta toplanabilir. Sınıflandırma metodu tahmin edici, kümeleme ve bağlantı analizi modeli ise tanımlayıcı metotlar olarak tanımlanabilir.

Çalışmamızın uygulama bölümünde sınıflandırma yöntemleri kullanılmış olması nedeniyle söz konusu öğrenme metoduna aşağıda detaylı olarak yer verilmekle birlikte, diğer makine öğrenmesi metotlarına yalnızca başlıklar halinde değinilmiştir.

2.2.1 Sınıflandırma

Veri sınıflandırması, öğrenme aşaması ve sınıflandırma aşaması olmak üzere 2 aşamalı bir süreçtir. İlk olarak, belirlenen veri kümesini tanımlayan bir sınıflandırıcı oluşturulur. Öğrenme aşaması (eğitim aşaması) olarak adlandırılan bu süreçte sınıflandırma algoritması, veri tabanı demetlerinden oluşan bir eğitim setini analiz ederek veya öğrenerek sınıflandırıcı oluşturur. Sınıflandırma aşamasında ise ilk olarak sınıflandırıcının başarı oranı tahmin edilir. Başarı oranının kabul edilebilir seviyede olması durumunda ise, sınıflandırıcı verinin diğer bölümü için kullanılabilir [17].

Sınıflandırma, bir veri kümesi üzerinde tanımlı olan muhtelif sınıflar arasında veriyi dağıtmak olarak tanımlanabilir. Bu yöntem için kullanılan algoritmalar, belirlenen eğitim kümesi içerisinde yer alan dağılım şeklini öğrenirler ve akabinde sınıfının belirsiz olduğu test verileri iletildiğinde uygun biçimde sınıflandırmayı hedefler [3].

Denetimli (supervised, gözetimli) bir öğrenme tekniği olan sınıflandırma metodunun temel özellikleri;

- verilerin etiketlerinin bulunması,
- verileri bir gruba dahil etmek için kural oluşturulması gerekliliği,
- veri setinin eğitim ve test olarak ayrılması gerekliliği

olarak belirtilebilir. Örnek vermek gerekirse, bir banka tarafından kullandırılan kredileri ödemeyen bir grup müşterinin verilerinden hareket ederek, yeni kredi talebinde bulunan müşterilerin kullanmak istedikleri krediyi ödeyip ödemeyeceklerini tahmin etmek için sınıflandırma metodu kullanılabilir.

Yaygın olarak kullanılan sınıflandırma modelleri;

- Lojistik Regresyon (Logistic Regression),
- Karar Ağaçları (Decision Trees),
- K-En Yakın Komşu (K-Nearest Neighbor),
- Yapay Sinir Ağları (Artificial Neural Networks),
- LightGBM

olarak sıralanabilir.

2.2.1.1 Lojistik Regresyon

Bir veya birden çok değişkenin diğer değişkenler cinsinden tahminini sağlayacak ilişkileri belirleme ve tanımlama işlemine regresyon analizi adı verilir. Bu analizde, gözlenen olayın değerlendirilmesi aşamasında bu olayın hangi olaylardan

etkilendiğinin tespit edilmesi önem taşımaktadır. Söz konusu olaylar bir veya birden çok olabileceği gibi etkileri farklı düzeylerde olabilir. Regresyon analizinde veriler arasındaki ilişki matematiksel bir fonksiyon olarak tanımlanır. Belirlenen matematiksel fonksiyon tek bağımlı değişken ve bir veya daha çok bağımsız değişkenden meydana gelebilir. Tek değişkenli modellere basit doğrusal regresyon, birden fazla bağımsız değişkenli modellere ise çoklu regresyon modelleri adı verilir [22].

“Y” bağımlı değişkeni ile yalnızca bir “X” bağımsız değişkeni arasındaki ilişkiyi doğrusal fonksiyonla ifade eden “basit regresyon modeli” aşağıdaki denklem ile ifade edilebilir.

$$Y = \alpha + \beta.X + \epsilon \quad (2.1)$$

Basit regresyon denkleminde yer alan;

- i. “ α ” katsayısı doğrusal fonksiyonun sabitini ifade eder. Bu katsayı, bağımsız değişken “X”in “0” değeri aldığı anda, bağımlı değişken “Y”nin regresyon doğrusundaki konumunu gösterir.
- ii. “ β ” katsayısı doğrusal fonksiyonun eğimini belirtir. Fonksiyonun eğimi, “X” bağımsız değişkeninde meydana gelen bir birimlik değişiminin, “Y” bağımlı değişkeni üzerinde yaratacağı değişimi gösteren regresyon katsayısıdır.
- iii. “ ϵ ” değeri ise fonksiyonun hata terimini ifade eder. Bu hata değeri, regresyon modeline net olarak dahil edilemeyen diğer değişkenleri ve durumları kapsamak üzere modele eklenen bir değerdir.

Bu durumda tahmini model;

$$\hat{Y} = \bar{\alpha} + \bar{\beta}.X + \epsilon \quad (2.2)$$

şekline dönüşür. “ α ”nın ve “ β ”nın tahminleri olan “ $\bar{\alpha}$ ” ve “ $\bar{\beta}$ ” değerleri genelde hata teriminin normal dağılım göstermesi durumunda en küçük kareler tekniğinden (EKK) veya hata teriminin dağılımı hakkında herhangi bir varsayım söz konusu değilse en büyük olasılırlık tekniklerinden yararlanılarak bulunulur. Uygulamada genelde EKK yöntemine başvurulur. EKK yönteminde amaç regresyon denkleminin hata değerini minimum kılacak parametre değerlerini belirlemektir. Regresyon fonksiyonunun hata değerinin minimum olmasını sağlamak için “ $\bar{\alpha}$ ” ve “ $\bar{\beta}$ ”

değerlerine göre “ ε ” hata değerinin kısmi türevlerinin alınarak “0”a eşitlenmesi gerekir. Bu işlem sonucunda elde edilen formüller kullanılarak “ $\bar{\alpha}$ ” ve “ $\bar{\beta}$ ” katsayı değerleri bulunabilir. Daha sonra regresyon modeli kullanılarak açıklayıcı değişkenin çeşitli değerlerine karşılık gelen bağımlı değişkenin alacağı değer tahmin edilebilir [2].

Çok değişkenli regresyon modelinde ise, bir bağımlı değişkenin yanı sıra birden fazla bağımsız değişken bulunmaktadır. Bu modelde bağımsız değişkenlerin bağımlı değişkendeki farklılıkları eş zamanlı olarak açıklaması hedeflenmektedir. Çok değişkenli regresyon modelinin denklemi ise aşağıdaki şekilde ifade edilebilir.

$$Y_1 = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \dots + \varepsilon_i \quad (2.3)$$

Lojistik regresyon modelinin temel amacı da, diğer regresyon modellerine benzer şekilde, bir veya birkaç bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi ifade etmektir. Lojistik regresyon yöntemi, diskriminant analizinde olduğu gibi doğrusallık ve sürekli veri zorunluluğu gibi kısıtlamalar içermediğinden kategorik değişken tahmini ve sınıflandırmasında tercih edilen bir yöntemdir. Ancak lojistik regresyon analizi neticesinde ulaşılan model temelde doğrusal olmayan bir logaritmik fonksiyondur ve ulaşılan katsayılar logaritmik değerleri gösterir. Söz konusu katsayıların anlamlı bir şekilde yorumlanması için odds oranından ve buna bağlı lojit değerlerinden faydalanılır. Odds oranı, (olabilirlik veya olasılık oranı) bir olayın gerçekleşme ihtimalinin $p(x)$, bu olayın gerçekleşmeme olasılığına “ $1-p(x)$ ” bölümü olarak tanımlanıp aşağıdaki şekilde gösterilir;

$$\text{Odds} = p(x)/1-p(x) \quad (2.4)$$

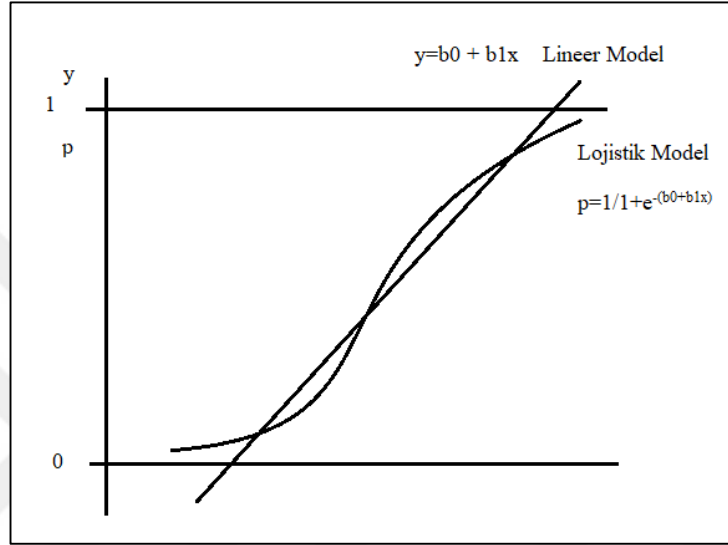
Odds oranı olasılık kestiriminin 0-1 arasında değer almasını sağlamakla birlikte, odds’un sıfırın altında bir değer almamasının da temin edilmesi gerekmektedir. Bunun için odds’un logaritması alınıp lojit değeri elde edilir. Lojit değeri aşağıdaki denklem kullanılarak hesaplanabilir;

$$\text{Ln } P(y=1)/P(y=0) = \text{Ln } (p/1-p) = e^{(\beta_0 + \beta_1 x_1)} \quad (2.5)$$

Ulaşılan lojit değerler ise üstel lojistik katsayıları verir. Bu katsayılar belirlendikten sonra aşağıdaki olasılık fonksiyonu kullanılarak yeni “x” gözleminin hangi sınıfa ait olacağı olasılık değeri olarak bulunabilir.

$$P=1/1+e^{-(b_0+b_1x_1+b_2x_2+\dots+b_px_p)} \quad (2.6)$$

Böylece lojistik regresyon, kategorik bağımlı değişken ile bağımsız değişkenler arasında ilişkiyi logaritmik dönüşümler yaparak doğrusal bir yapıya dönüştürür ve her bir gözlem için belli sonuçların gerçekleşme olasılığını belirler. Aşağıdaki grafikte bu dönüşüm gösterilmiştir.



Şekil 4: Lojistik Regresyon Fonksiyonu [2]

Gerek akademik araştırmalarda gerekse iş analizlerinde elde edilen verilerin önemli bölümü kategorik verilerdir. Örnek vermek gerekirse, bir kişi bir akademik programı ya tamamlamıştır ya da tamamlamamıştır, benzer şekilde bir kişinin araç sürme ehliyeti vardır ya da yoktur. Bu nedenle, kategorik verilerin çok değişkenli analizi birçok sektörün ilgi alanına girmektedir. Dolayısıyla lojistik regresyon analizi, regresyon analizi temeliyle ve diğer analizlere göre daha yaygın bir kullanım alanı sağlamasının avantajıyla, kategorik verilere ilişkin çalışmalarda büyük öneme sahiptir.

2.2.1.2 Karar Ağaçları

Karar ağaçları, sınıfları belli olan örnek veriden tümevarım yoluyla öğrenilen ağaç şeklinde bir karar yapısı türüdür. Bir karar ağacı, basit karar verme adımları izlenerek, büyük miktarlardaki kayıtları küçük kayıt gruplarına ayırarak kullanılan bir yapıdır. Her başarılı ayırma işlemi sonrasında, sonuç gruplarında yer alan üyelerin diğer üyelere benzerliği artmaktadır [23].

Karar ağaçları yönteminde öncelikle sınıflandırma amacıyla bir ağaç oluşturularak, veri tabanında yer alan kayıtlar bu ağaca uygulanır ve elde edilen sonuca göre kayıt sınıflandırılır. Uygulamanın yanı sıra, sonuçların anlaşılması ve yorumlanması aşamalarının daha kolay olması nedeniyle karar ağaçları diğer yöntemlere kıyasla yaygın olarak kullanılmaktadır. Genellikle kategorik niteliklerin kullanıldığı sınıflandırmada, sürekli sayısal değişken varsa bu veriler kategorik şekle dönüştürülerek karar ağacında kullanılabilir. Karar ağacı algoritmasının ağacın öğrenilmesi aşamasında eğitim amaçlı kullanılan küme, çeşitli niteliklere göre alt kümelerle ayrılır. Söz konusu işlem, öz yineli olarak tekrarlanır ve tekrarlama işleminin tahmin üzerinde etkisi kalmayana kadar devam eder [3].

Karar ağaçları kurulum maliyetinin düşük olması, kolay yorumlanabilmesi, veri tabanı sistemleri ile rahatlıkla bütünleştirilebilmelerinin yanı sıra, tatmin edici seviyede güvenilirlikleri bulunması nedenleriyle yaygın kullanıma sahiptir [24].

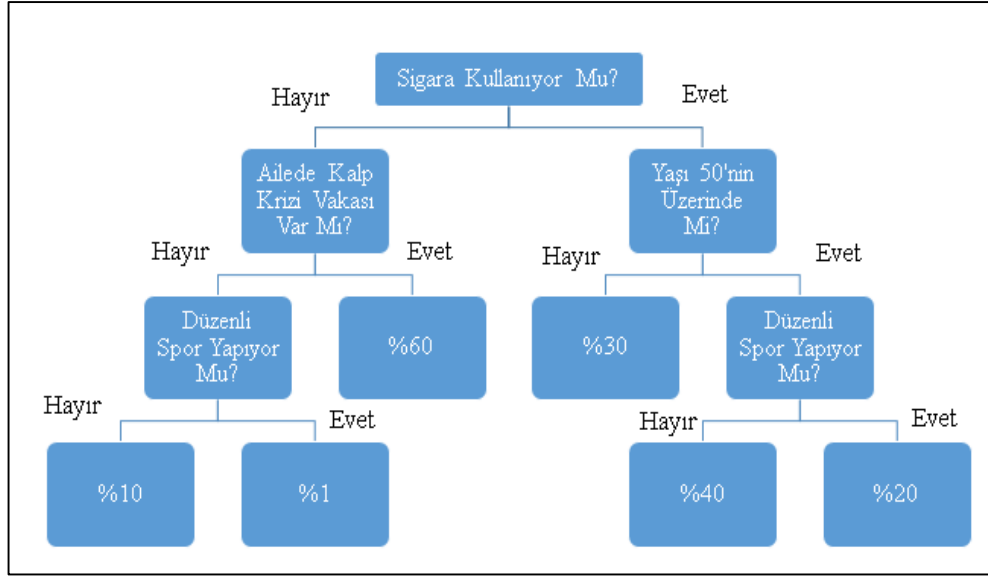
Karar ağaçları oluşturulurken kullanılan algoritma seçimi son derece önem taşımaktadır. Tercih edilen algoritmaya göre karar ağacının şekli değişiklik gösterebilir. Değişik ağaç yapıları farklı sınıflandırma sonuçları verecektir. Kök adı verilen ilk düğümün farklı olması, uçtaki yaprağa ulaşılırken takip edilecek yolu ve bu sebeple sınıflandırmayı değiştirecektir. Karar ağacı algoritmasında amaç, en kısa yoldan istenen yanıtı ya da sınıfa ulaşmaktır [7].

Farklı yaklaşımlar ile oluşturulan karar ağacı uygulamalarından yaygın olarak kullanılanları;

- CHAID,
- CRT,
- ID3,
- C4.5,
- MARS

modelleri olarak ifade edilebilir.

Aşağıdaki şekilde temsili bir veri setinde yer alan kişilerin kalp krizi yaşaması ihtimalini gösteren örnek bir karar ağacına yer verilmiştir.



Şekil 5: Örnek Karar Ağacı

Yukarıdaki karar ağacında yer verilen her düğüm kendisinden sonra iki dala ayrılmaktadır. Değerilen ayrılma işlemi sürecinde yeni düğüm hakkında cevabı veri tabanında bulunacak bir soru sorulmakta ve verilecek yanıtla göre bir dal izlenmektedir. Yukarıdaki şekilde yer verilen kayıtların yorumlanması sonucunda;

- sigara kullanan, 50 yaşının üzerinde olan ve düzenli spor yapmayan bir kişinin kalp krizi geçirmesi ihtimali %40,
- sigara kullanmayan, ailesinde kalp krizi vakası bulunmayan ve düzenli spor yapan bir kişinin kalp krizi geçirmesi ihtimali %1

olarak hesaplanacaktır.

Karar ağacı algoritmalarının öne çıkan avantajları, parametrik olmayan yöntemlerden olması sebebiyle diğer çok değişkenli tekniklerde sağlanması gereken istatistik varsayımların söz konusu olmamasıdır. Çünkü veri madenciliği sonucunda oluşturulan kuralların kullanılıp kullanılmayacağı yönünde bir yargıya önceden varılması mümkün değildir. İlaveten karar ağacının bir diğer avantajı, algoritmalarının bağımlı ve bağımsız değişkenler arasındaki ilişkilerin yönünü ve önem sırasını görselleştirmesidir. Bu sayede ulaşılan sonuçların yorumu basitleştirilerek daha somut ve kullanışlı hale getirilmektedir [23].

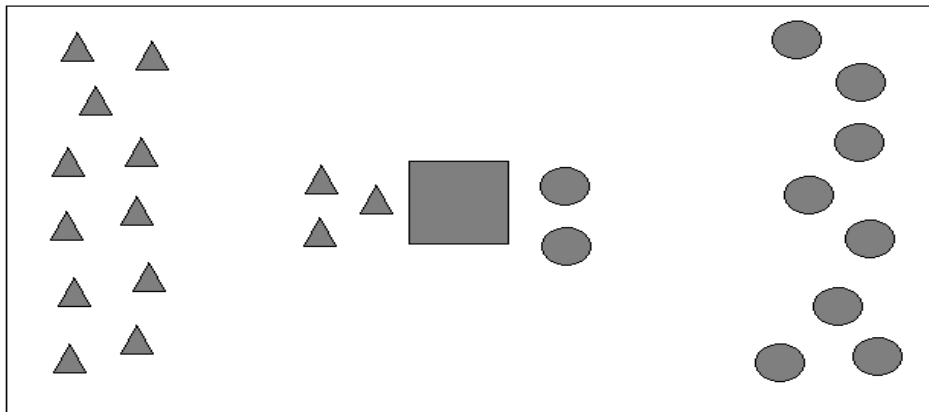
2.2.1.3 K-En Yakın Komşu (K-NN)

K-En Yakın Komşu algoritması sınıfları belli olan bir örnek kümesinde yer alan gözlem değerlerinden faydalananarak örneğe dahil edilecek yeni gözlemin hangi

sınıfa ait olduğunu belirlemek için kullanılır. K-En Yakın Komşu yöntemi, örnek kümedeki gözlemlerin her birinin sonradan belirlenen bir gözlem değerine uzaklıklarının hesaplanması ve en küçük uzaklığa sahip “k” sayıda gözlemin seçilmesi temeline dayanmaktadır. Uzaklıklar, Öklid, Minkowski, Manhattan, Chebyshev veya Dilca formülleri kullanılarak hesaplanabilir. K-En Yakın Komşu algoritması kullanılırken, gözlem değerlerinden meydana gelen küme için aşağıda yer verilen aşamalar izlenir [25]:

- i. K değeri belirlenir. Bu değer belirlenen bir noktaya en yakın mesafedeki komşuların sayısıdır.
- ii. Algoritma belirlenen bir noktaya en yakın komşuları tespit edeceği için, söz konusu nokta ile diğer tüm noktalar arasındaki uzaklıklar tek tek hesaplanır.
- iii. Önceki adımda hesaplanan uzaklıklar dikkate alınarak satırlar sıralanır ve bu satırlar içerisinde en küçük olan “k” tanesi seçilir.
- iv. Belirlenen satırların ait olduğu kategori tespit edilir ve en çok yinelenen kategori değeri seçilir.
- v. Seçilen kategori, tahmin edilmesi beklenen gözlem değerinin kategorisi olarak dikkate alınır.

Bir örnekle açıklamak gerekirse, aşağıdaki şekilde yer verilen kare şeklindeki simgenin, sağ tarafında yer alan yuvarlak simgelere mi yoksa sol tarafında yer alan üçgen simgelere mi dahil olacağı kararı “k” değerine göre değişecektir. Şöyle ki, “k=2” olması durumunda kare nesneye en yakın 2 nesne sağ tarafında yer alan yuvarlaklar olacağı için, kare nesne yuvarlak grupta sınıflandırılacaktır. Diğer taraftan, “k=3” olması durumunda ise, kare nesneye en yakın 3 nesne sol tarafında yer üçgenler olacağı için, kare nesne üçgen grupta sınıflandırılacaktır.



Şekil 6: K-NN Algoritması Örneği

K-NN algoritması; eğitim sürecinin hızlı olması, yöntemin öğrenilmesinin yalın ve kolay olması, eğitim için ayrılan verinin büyük olması durumunda efektif bir yöntem olması ve gürültülü eğitim verilerine karşı dirençli olması gibi avantajlar ile öne çıkmaktadır. Diğer taraftan, performansının “k” komşu sayısına bağlı olarak etkilenmesi, yüksek miktarda bellek alanına gereksinim duyması, ilgili olmayan veri nedeniyle yanlışma ihtimalinin bulunması, muhtelif hesaplama karmaşıklıkları barındırması nedenleriyle ise bazı dezavantajlar barındırmaktadır [26].

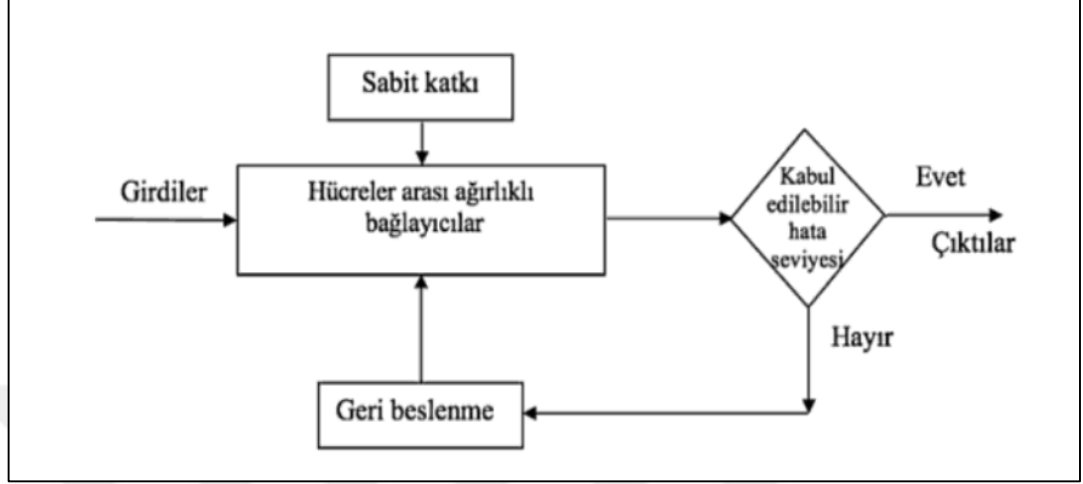
2.2.1.4 Yapay Sinir Ağları

Yapay sinir ağları modeli, nöronlardan oluşan biyolojik sistemlerin davranışına benzer bir yapı ortaya koymayı amaçlamaktadır. 1950’li yıllardan bu yana yapay sinir ağları hem sınıflandırma hem de sürekli regresyon modellerinin tahmini amacıyla kullanılmaktadır. Yapay sinir ağları modeli bilhassa, sınıf değerleri ve özellikleri arasındaki ilişkinin net olarak bilinmediği veya girdi ve çıktılar arasındaki ilişkinin doğrusal olmadığı durumlarda tercih edilen bir yöntemdir [2].

Yapay sinir ağları, yapay sinir hücrelerinin birbirleriyle muhtelif şekillerde bağlanmasından meydana gelir ve genel itibarıyla katmanlar şeklinde düzenlenir. Öne çıkan özellikleri birbirlerine bağlı nöronlar, bağlantılar arasındaki ağırlıkların tespit edilmesi ve ateşleme fonksiyonudur. Yapay sinir ağlarını meydana getiren her nöronun bir iç hali bulunmaktadır. Bu iç hale aktivasyon veya aktivasyon seviyesi adı verilir. Bu seviye, gelen girdileri tanımlayan bir fonksiyondur. Ağ içerisindeki bir nöron, doğal nöronların yaptığına benzer biçimde, diğer nöronlara bir işaret (sinyal) gönderir. Nöronlar bir seferde tek bir işaret gönderebilirler. İletilen bu sinyaller gönderilen nöronlar için giriş fonksiyonları olur. Bir nöron bir kerede birden fazla nörona sinyal gönderebilir [7].

Yapay sinir ağları modelinde bilgilerin işlenmesi eş zamanlı gerçekleştirildiğinden iletilen bilgiler birbirinden bağımsızdır. İlaveten, aynı tabakada yer alan bağlantılar arasında zaman bağımlılığı bulunmaması nedeniyle tamamı ile eş zamanlı çalışabilmekte ve bu sayede bilgi akış hızı artmaktadır [27]. Yapay sinir ağları modelinde öğrenme, bağlantı ağırlıklarının yenilenmesi şeklinde meydana gelmektedir. Öğrenme neticesinde ulaşılan bilgiler bağlantı ağırlıklarında saklanır. Bu sayede ulaşılan bilginin uzun süre muhafaza edilmesi mümkündür. Ayrıca öğrenme yeteneği tam tanımlı olmayan problemlerin yapay sinir ağları ile çözülmesine imkan sağlamaktadır. Paralel çalışan bir yapay sinir ağı modeli karışık fonksiyonlarla yapılan

işlemler yerine basit işlemler içerdiğinden ve karmaşık olmayan bir mimari yapıya sahip olduğundan birçok sorunun çözülmesinde tercih sebebidir. Aşağıdaki şekilde yapay sinir ağı modelinin genel yapısına yer verilmiştir.



Şekil 7: Yapay Sinir Ağı Modeli Şeması [28]

Daha önce değinildiği üzere, yapay sinir ağları algoritması biyolojik sinir sistemi örnek alınarak tasarlanmıştır. Biyolojik sinir hücreleri arasındaki iletişim synapsler vasıtasıyla sağlanır. Bir sinir hücresinin işlediği bilgiler axon vasıtasıyla ile diğer hücrelere iletir. Benzer biçimde yapay sinir hücreleri dışarıdan gelen bilgileri toplama fonksiyonu ile toplar ve aktivasyon fonksiyonundan geçirerek çıktıyı üretir, ağıın bağlantıları üzerinden diğer hücrelere (proses elemanlarına) gönderir. Değişik toplama ve aktivasyon fonksiyonları bulunmaktadır. Yapay sinir ağlarını birbirlerine bağlayan bağlantıların değerleri ağırlık değerleri olarak adlandırılır. Proses elemanları paralel üç katman biçiminde birleşerek bir ağ meydana getirirler. Bu katmanlar; girdi katmanı, ara katman ve çıktı katmanı olarak belirtilebilir. Ağa girdi katmanı üzerinden dahil edilen bilgi, akabinde ara katmanda işlenir ve çıktı katmanına aktarılır. Bilgi işleme olarak belirtilen husus, ağa erişen bilgilerin ağıın ağırlık değerleri kullanılarak çıktıya çevirilmesidir. Doğru çıktılara erişilebilmesi için ağıın ağırlıklarının yanlış değerlerden oluşmaması önemlidir. Doğru ağırlıkların bulunması işlemi ağıın eğitilmesi olarak ifade edilmektedir. Söz konusu değerler ilk etapta rasgele atanırlar. Takip eden süreçte, eğitim aşamasında her örnek ağa iletildiğinde, ağıın öğrenme kuralı göz önünde bulundurularak ağırlıklar güncellenir. Sonrasında, farklı bir örnek ağa dahil edilerek ağırlıklar tekrar güncellenir ve en doğru değerler elde edilmeye çalışılır. Bu süreç ağ eğitimi setinde yer alan örneklerin tamamı için doğru çıktılar elde

edilinceye kadar tekrarlanır. Bu sağlandıktan sonra test setinde yer alan örnekler ağa gösterilir. Eğer ağ test setinde yer alan örneklere doğru cevaplar verirse ağ eğitilmiş sayılmaktadır [29].

2.2.1.5 LightGBM (Light Gradient Boosting Machine)

LightGBM (Light Gradient Boosting Machine), hızlı ve yüksek performanslı bir makine öğrenimi algoritmasıdır. Microsoft tarafından geliştirilen ve açık kaynaklı bir proje olan LightGBM, özellikle büyük veri kümeleri üzerinde çalışırken etkili sonuçlar elde edebilmek için tasarlanmıştır. LightGBM hızlı öğrenebilmesi, düşük bellek tüketimi ve yüksek tahmin performansı gibi özellikleriyle bilinmektedir.

LightGBM, bir makine öğrenmesi algoritması olan ve zayıf tahmincilerin (genellikle karar ağaçları) bir araya getirilerek daha güçlü bir tahminci oluşturulmasına dayanan gradient boosting yönteminin bir çerçevesidir.

LightGBM'in temel özellikleri şunlardır;

- Düşük Bellek Tüketimi: LightGBM, diğer geleneksel gradient boosting yöntemlerine göre daha az bellek kullanır. Bu sayede daha büyük veri kümeleri üzerinde çalışırken daha az bellek tüketimi sağlar.
- Hızlı Öğrenebilme Yeteneği: LightGBM, diğer gradient boosting algoritmalarına göre daha hızlı bir öğrenme yeteneğine sahiptir. Bu hızlı öğrenme, daha kısa eğitim süreleri anlamına gelir.
- Dengesiz Veri Kümesi Desteği: LightGBM, sınıf dengesizliği içeren veri kümeleri üzerinde iyi performans gösterir. Sınıf ağırlıklandırma ve örnekleme stratejileri ile bu tür sorunlara çözüm sunar.
- Histogram Tabanlı Öğrenme: LightGBM, veri setini histogramlara böler ve bu histogramlar üzerinden öğrenme işlemini gerçekleştirir. Bu sayede hızlı ve verimli hesaplamalar yapılabilir.
- Kategori Nitelik Desteği: LightGBM, kategori (kategori) niteliklerini özel olarak işleyerek hızlı bir şekilde kodlar ve böylece kategori niteliklerinin kullanılmasını kolaylaştırır.
- Dağıtık ve Paralel Öğrenme Desteği: LightGBM, birden fazla işlemci veya düğüm üzerinde dağıtık ve paralel öğrenme yapabilme yeteneğine sahip olup, bu sayede eğitim sürelerini daha da hızlandırır [35].

2.2.2 Kümeleme

Denetimsiz öğrenme tekniği (unsupervised) olan kümeleme metodunda, verilerin birbirlerine benzerlikleri dikkate alınarak gruplandırma yapılır. Marketlerde müşteri gruplarının belirlenmesi ve alışveriş örüntülerinin çıkarılması, bir ilçede bulunan konutların muhtelif özelliklerine göre ayrıştırılması vb. uygulamalar kümelemeye örnek olarak gösterilebilir.

Verilerin etiketinin bulunmadığı kümeleme metodunda, veri yakınlığına, benzerliğine ve alakasına göre bir gruba dahil edilir. Yaygın kullanılan kümeleme modelleri;

- K-Means Kümeleme Algoritması,
- Hiyerarşik Yöntemler,
- Yoğunluk Tabanlı Yöntemler,
- Model Tabanlı Yöntemler

olarak belirtilebilir.

2.2.3 Bağlantı Analizi Modeli (Birliktelik Kuralları)

Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri tespit eder. Elde edilen ve depolanan verinin gün geçtikçe büyümesi nedeniyle, işletmeler veri tabanlarında yer alan birliktelik kurallarını tespit etmeye önem vermektedir. Büyük miktardaki mesleki işlem verilerinden muhtelif birliktelik ilişkilerini tespit etmek, işletmelerin karar alma süreçlerindeki verimliliğini artırmaktadır.

Birliktelik kurallarının yaygın olarak uygulandığı örnek market alışveriş uygulamasıdır. Bu uygulama, satın aldıkları ürünler arasındaki ilişkiyi tespit ederek müşterilerin alışveriş alışkanlıklarını analiz etmektedir. Bu tip birlikteliklerin belirlenmesi, müşterilerin bir arada aldıkları ürünleri ortaya çıkarır ve işletmeler bu bilgiyi kullanarak etkin satış stratejileri geliştirebilirler. Örnek olarak, bir müşterinin çay satın alması durumunda aynı alışverişte şeker alma olasılığının değerlendirilmesi vb. analizler gösterilebilir [30].

BÖLÜM III

BANKACILIKLA İLGİLİ TEMEL KAVRAMLAR

Bu çalışmada uygulama bankacılık sektöründe gerçekleştirilmektedir. Bu nedenle bu bölümde uygulamanın yapıldığı sektöre ilişkin temel bilgiler yer almaktadır. Bankacılık sektörüne ilişkin en temel düzenleme sektöre ilişkin çıkarılmış olan 5411 sayılı Bankacılık Kanunudur. 5411 numaralı Bankacılık Kanunu'nda yer alan tanımına göre bankalar;

- mevduat bankası; kendi nam ve hesabına mevduat kabul etmek ve kredi kullandırmak esas olmak üzere faaliyet gösteren kuruluşlar ile yurt dışında kurulu bu nitelikteki kuruluşların Türkiye'deki şubeleri,
- katılım bankası; özel cari ve katılma hesapları yoluyla fon toplamak ve kredi kullandırmak esas olmak üzere faaliyet gösteren kuruluşlar ile yurt dışında kurulu bu nitelikteki kuruluşların Türkiye'deki şubeleri,
- kalkınma ve yatırım bankası; mevduat veya katılım fonu kabul etme dışında; kredi kullandırmak esas olmak üzere faaliyet gösteren ve/veya özel kanunlarla kendilerine verilen görevleri yerine getiren kuruluşlar ile yurt dışında kurulu bu nitelikteki kuruluşların Türkiye'deki şubeleri

olarak tanımlanmıştır.

Bankacılık Kanunu'nda bankaların;

- mevduat ve katılım fonu toplayabileceği,
- kredi kullandırabileceği,
- ödeme, fon transferi ve tahsilat işlemleri yapabileceği,
- ilaveten, kanunun 4. maddesinde yer verilen diğer faaliyetleri yürütebileceği

hüküm altına alınmıştır. 2023 yılının Haziran ayı itibarıyla ülkemizde 35 adedi mevduat bankası, 19 adedi kalkınma ve yatırım bankası, 6 adedi ise katılım bankası olmak üzere toplam 60 banka; 11.048 şube ve 207.266 personel ile faaliyet göstermektedir [31]. Aynı zamanda 6 adet dijital banka kuruluş izni almış olup, faaliyete geçme konusunda çalışmaları devam etmektedir.

Genel bir değerlendirme yapmak gerekirse bankalar, gerçek ve tüzel kişilerin tasarruflarını toplayarak, bunları gelir sağlayacak işlere kredi ve benzeri yollarla yönlendiren, ödemelerde aracılık yapan, para nakli ve kıymetli evrak tahsili, emanet kabulü gibi çeşitli hizmetler gören işletmelerdir. Diğer bir tanımla bankalar; özellikle var olan öz kaynaklarından ve müşterilerinden toplanan mevduatlardan ortaya çıkarılan olanakları kullanarak para ve kredi ticareti yapan, ayrıca bunların dışındaki finansal ve toplumsal faaliyetlere girerek hayata katkıda bulunan işletmeler olarak ifade edilebilir. Değerlenen tanımlardaki unsurlardan yola çıkarak, bankacılık faaliyetlerinin kaynak sağlamak ve sağlanan kaynakları fon ihtiyacı olan kişi ve kurumlara sunmak olacak şekilde iki temel fonksiyonu bulunduğu sonucuna ulaşılabilir [32].

3.1 BANKACILIKTA KREDİ KAVRAMI

Kredi, bankanın yapacağı istihbarat neticesinde, gerçek veya tüzel kişilere yasal mevzuat, kurum içi politikalar ve kaynaklar göz önünde bulundurularak teminat karşılığı ve/veya teminatsız olarak para, teminat veya kefalet vermek şeklinde tanınmış olan olanak veya limit olarak tanımlanabilir [33]. 5411 numaralı Bankacılık Kanunu'nda; bankalar tarafından kullanılan nakdi ve gayrinakdi krediler ile bu niteliğe sahip taahhütler ve ilgili Kanunun 48. Maddesinde yer verilen diğer işlemler kredi olarak nitelendirilmiştir.

Bankacılık Düzenleme ve Denetleme Kurulu verilerine göre, ülkemizde bankacılık sektörünün 2023 yılının Haziran ayına ilişkin özet verilerine aşağıdaki tabloda yer verilmiştir.

Tablo 1: Bankacılık Sektörü Bilançosu (Haziran/2023)

<u>BİLANÇO HESABI</u>	<u>TUTAR</u> <u>(milyon TL)</u>	<u>BİLANÇO HESABI</u>	<u>TUTAR</u> <u>(milyon TL)</u>
Krediler	10.009.615	Mevduat (Katılım Fonu)	9.147.150
Zorunlu Karşılıklar	1.191.949	Karşılıklar*	209.927
Takipteki Alacaklar	168.306	Ödenmiş Sermaye	255.296
Beklenen Zarar Karşılıkları(-)	414.339	Dönem Kârı	250.136
Diğer Aktifler	7.318.180	Diğer Pasifler	9.239.879
TOPLAM AKTİFLER	19.102.388	TOPLAM PASİFLER	19.102.388

* TFRS 9 uygulamayan bankalar tarafından ayrılan "Genel Karşılık" tutarını da kapsamaktadır.

Detaylarına yukarıdaki tabloda yer verildiği üzere, 2023 yılının Haziran ayı itibarıyla Türkiye bankacılık sektörü bilançosunun aktifinin %52,4'ü kredi kalemlerinden oluşmaktadır.

3.2 BANKACILIKTA TEMERRÜT VE KARŞILIK KAVRAMLARI

Temerrüt; taraflar arasındaki borç ilişkisine sebep olan borcun ya da yükümlülüğün sözleşmede belirtilen sürede yerine getirilmemesi durumudur. Bir diğer ifadeyle borçlu olan taraf edimini zamanında ifa edememesi durumunda temerrüde düşer. Karşılık ise; kredi zararlarının karşılanması amacıyla mali tablolarda hesaben ayrılan ve gider yazılan tutarlar anlamına gelmektedir. Geri ödemelerinde yaşanan aksaklıklar göz önünde bulundurularak krediler farklı gruplar içerisinde izlenmekte olup, ilgili gruplara denk gelen karşılık oranına tabi tutulur.

Bankaların ve diğer finans kuruluşlarının sürekliliğini sağlayabilmesi amacıyla toplayacağı mevduatı kredi olarak talep eden kişilere aktarması ve bu sayede faaliyetine devam etmesi gerekmektedir. Bu nedenle, verilen kredilerin mümkün olduğu ölçüde temerrüde düşmemesi, bu sayede ayrılacak karşılık miktarının azaltılması ve kredi geri ödenmelerinin sağlanarak yeni kredilerin kullanılması esastır.

Kredilerin niteliklerine göre sınıflandırılması ve bunlar için ayrılacak karşılıklara ilişkin usul ve esaslar, BDDK tarafından yayınlanan “Kredilerin Sınıflandırılması ve Bunlar İçin Ayrılacak Karşılıklara İlişkin Usul ve Esaslar Hakkında Yönetmelik” ile hüküm altına alınmıştır. Değinilen yönetmeliğe göre krediler aşağıdaki şekilde 5 farklı grupta izlenmektedir;

a. Birinci Grup (Standart Nitelikli) Krediler içerisinde;

- i. Kredilendirmeye uygun mali yapıyı haiz kişilere kullanılan,
- ii. Anapara ve faiz ödemeleri borçlusunun ödeme imkanına ve nakit akışı dikkate alınarak düzenlenen,
- iii. Ödemeleri vadesinde gerçekleştirilen veya otuz günden fazla gecikmesi bulunmayan, ilerleyen dönemlerde de ödemelerinin düzenli olması beklenen, teminatlara başvurulmadan tortusuz biçimde tahsil imkanı olan,
- iv. Borçlusunun kredi değerliliği zayıflamamış olan,
- v. İlgili yönetmeliğin 4. maddesinin (a) başlığında yer verilen diğer koşulları taşıyan

krediler yer alır.

b. İkinci Grup (Yakın İzlemedeki) Krediler başlığında;

- i. Kredilendirmeye uygun mali yapıyı haiz olan kişilere tahsis edilen fakat ekonomik koşullarda veya borçlunun faaliyet gösterdiği sektördeki olumsuzlukların yanı sıra, borçlunun muhtelif olumsuzluklar sebebiyle ödeme kabiliyetinde veya nakit akışında menfi gelişmeler tespit edilen ya da bunun gerçekleşeceği öngörülen,
- ii. Tahsis edildiği dönemde borçlunun kayda değer seviyede finansal risk barındırması vb. sebeplerle yakından izlenmesi uygun olan,
- iii. Anapara/faiz ödemelerinin üzerinde anlaşılan koşullara uygun biçimde gerçekleştirilmemesi ihtimali bulunan ve buna sebep olan unsurların ortadan kaldırılamaması durumunda teminatlara başvurulmadan borcun tortusuz biçimde ödenememe ihtimali barındıran,
- iv. Kredi kullandırıldığı tarihe kıyasla borçlusunun kredi değerliliği zayıflamamış olmasına rağmen, nakit akışındaki düzensizlik sebebiyle zayıflama riski bulunan,
- v. Kredi değerliliğinin zayıflaması olarak değerlendirilemeyecek sebeplerle anapara/faiz ödemelerinin sözleşmelerde belirlenen tarihlerden itibaren otuz günden fazla geciken ancak doksan günü aşmayan,
- vi. İlgili yönetmeliğin 4. maddesinin (b) başlığında yer verilen diğer unsurları haiz

krediler yer almaktadır.

c. Üçüncü Grup (Tahsil İmkânı Sınırlı) Krediler:

- i. Borçlusunun kredi değerliliği bozulmuş olan,
- ii. Teminatların net gerçekleşebilir değerinin veya borçlunun özkaynaklarının borcun vadesinde ödenmesini karşılayacak seviyede olmaması sebebiyle teminata başvurulmadan tamamının tahsil imkanı kısıtlı bulunan ve gözlenen sorunların giderilmemesi durumunda zarara yol açması muhtemel görülen,
- iii. Anaparanın/faizin tahsili, vadesini veya sözleşmede belirlenen tarihi doksan günden fazla geçen fakat yüz seksen günü geçmeyen,
- iv. İlgili yönetmeliğin 4. Maddesinin (c) başlığında yer verilen diğer unsurlara sahip

krediler ve diğer alacaklar sınıflandırılır.

ç. Dördüncü Grup (Tahsili Şüpheli) Krediler;

- i. Anaparanın/faizin tamamının kredi sözleşmesinde yer alan koşullar kapsamında teminata başvurulmadan tahsil edilemeyeceği muhtemel bulunan,
- ii. Anapara/faizin vadesinden veya ödenmesi gereken tarihten itibaren tahsilinin gecikmesi yüz seksen günü geçen ancak bir yılı geçmeyen,
- iii. İlgili yönetmeliğin 4. Maddesinin (ç) başlığında yer verilen diğer unsurlara sahip

krediler sınıflandırılır.

d. Beşinci Grup (Zarar Niteliğindeki) Krediler;

- i. Borçlusunun kredi değerliliğini yitirmesi nedeniyle tahsil edilemeyeceği öngörülen veya toplam alacak tutarının yalnızca ihmal edilebilir bir bölümünün tahsil edilebileceği görüşüne ulaşılan,
- ii. Üçüncü ve Dördüncü Gruplarda yer verilen özelliklere sahip olmakla beraber, sözleşmede yer alan alacak tutarlarının tamamının bir yıldan daha uzun bir sürede tahsili mümkün görünmeyen,
- iii. İlgili yönetmeliğin 4. Maddesinin (d) başlığında yer verilen diğer unsurlara sahip

krediler sınıflandırılır. Aynı yönetmelik kapsamında Üçüncü, Dördüncü, Beşinci grupta izlenen ve tahsil imkanı düşük bulunan krediler “**donuk alacak**” olarak nitelendirilmektedir.

Kullandırılan kredilerin tamamının geri ödemelerinin sözleşmeye uygun biçimde yapılmasının uygulamada çok mümkün olmadığı ve belirli bir kısmının temerrüde düşerek donuk alacak olarak muhasebeleştirileceği tabiidir. Bu aşamada, kullandırılması durumunda ilerleyen dönemde donuk alacak olarak gruplandırılması muhtemel kredilerin tahsis aşamasında mümkün olduğu ölçüde öngörülmesi, bu doğrultuda kredinin kullandırılmaması ve bankanın mali bünyesinin olumsuz etkilenmemesi önem kazanmaktadır.

Finansal işletmeler olarak sermaye yapılarını kuvvetli tutmaları gereği haricinde bankaların, Bankacılık kanunu da dahil olmak üzere, sermayelerinin belirli kriterleri sağlamasına yönelik olarak bir takım yasal düzenlemelere uyması zorunludur. 5411 sayılı Bankacılık Kanununda; “Bu kanunun uygulanmasında maruz kalınan riskler nedeniyle oluşabilecek zararlara karşı yeterli özkaynak bulundurulması

sermaye yeterliliğini ifade eder. Bankalar, Bankacılık Düzenleme ve Denetleme Kurulu tarafından düzenlenecek yönetmelikte öngörülen usul ve esaslara göre yüzde sekiz oranından az olmamak üzere belirlenecek sermaye yeterliliği oranını hesaplamak, tutturmak, idame ettirmek ve raporlamak zorundadır.” hükmü bulunmaktadır. Dolayısıyla bankaların, faaliyetlerini sağlıklı bir biçimde sürdürebilmeleri için aktif kalitelerini artırmaları, kuvvetli bir sermaye yapısına sahip olmaları ve bunun için geri ödenmeyen kredilerin toplam krediler içerisindeki payını mümkün olduğu ölçüde düşük seviyede tutmaları gerekmektedir.

Konu ile ilgili olarak muhtelif görüşlere aşağıda ayrıca yer vermiştir;

- Bankacılık sisteminin temel araçları arasında yer alan kredi, kısmen veya tamamen geri ödemesi belirli süreyi geçtiğinde (çalışmanın yapıldığı tarih itibarıyla 90 gün) takipteki kredi olarak dikkate alınır. Takipteki kredilerin toplam kredi içerisindeki payı, ekonomide kişi veya işletmelerin ödeme potansiyelini, bankalarda ise aktif kalitesini ve risk seviyesini gösterir. Kredilerin takibe intikal etmesinin, ekonomiye ve bankalara çok sayıda maliyeti vardır. Takip hesaplarında izlenen kredilerin payı, öncü gösterge olması nedeniyle ekonomi ve banka yönetimlerince yakından izlenir. Değinen sebeple, kullanılan kredilerin takip hesaplarına aktarılacağı öngörülmesi, bankalar ve ekonomi yönetimlerince önem atfedilen bir konudur. Takip hesaplarında izlenen kredilerin tahminine ilişkin gözlemlere ve ekonometrik modellere dayanan çeşitli sistemler bulunmaktadır. Sistemlerin önemli bir bölümü firma verilerini girdi olarak kullanırken, kalan kısmı bu verilerin yanı sıra makro ekonomik unsurları da modellerine dahil eder. Kâr artırmayı ve riski asgari seviyeye indirmeyi amaçlayan bankaların, kredilerin takip hesaplarına aktarılması ihtimalini ilgili krediyi kullandırma aşamasında sağlıklı biçimde tahmin edebilmeleri önem arz etmektedir [36].
- Bankaların asli görevi, fon arz edenler ile fon talep edenler arasındaki aracılık fonksiyonudur. Bu nedenle, bankalar fonların miktar, zaman, faiz, tarih açısından farklılıklarını uyumlaştırarak bu fonksiyonlarını etkin biçimde yerine getirmeye çalışmaktadır. Ayrıca bankalar, sahip oldukları bilgi, birikim ve muhtelif olanakları kullanarak fon arz edenlerin bireysel olarak yapamayacakları veya karar veremeyecekleri faaliyetleri yapma imkanına sahiptir. Bu işlemleri kâr amacı ile yapan bankalar, güven

müesseseleri olmalarının bir sonucu olarak bu süreci sağlıklı bir şekilde ve en az riskle yürütmeye gayret göstermektedirler. Öte yandan, kredi tahsis sürecinde asimetrik bilginin var olması, süreçte yapılacak hatalar, ekonomik olumsuzluklar, reel sektör işletmelerinden kaynaklanacak muhtelif olumsuzluklardan kaynaklı olarak kullandırılan kredilerin tahsil edilmesinde sorunlar yaşanabilmektedir. Sorunlu krediyi, banka ile borçlu arasındaki geri ödeme anlaşmasına uyulmaması nedeniyle tahsilatın zamanında yapılmaması ve zarar olasılığının ortaya çıkması şeklinde ifade edebiliriz. Kredilerin sınıflandırılması, bu süreçte yapılacak hatalı sınıflandırmaların banka mali tablolarını ve dolayısıyla kâr, sermaye yeterliği, aktif kalitesi vs. oranları etkilemesi ve bankaların sonuç rakamlarının gerçekten farklı görünmesine sebebiyet verebilmesi nedenleriyle, azami önem taşımaktadır. Bu durum denetleyici kurumlar, pay sahipleri, banka müşterileri gibi banka sonuç rakamlarını analiz eden kişilere yanıltıcı bilgi verilmesine yol açabilir. Sorunlu krediler banka ve reel sektör firmalarına etkileri sebebiyle, bankaların önem atfettikleri ve tahsili için yakından takip ettikleri bir süreçtir [37].

Çalışmamızın hazırlandığı dönem itibarıyla kredi başvurularında ve tüketimde yaşanan artışa bağlı olarak geri ödemeleri sözleşmelerde belirlenen şartlara uygun olarak yapılmayan banka kredilerinde artış görülmektedir. Bankaların veya finansal kuruluşların, talepte bulunan kişinin alışkanlıklarını, para harcama alışkanlıklarını, temerrüt tahminlerini ve diğer çeşitli modern tahmin model teknikleri kullanarak başvuru sahibi hakkında öngörü elde edebilmeleri mümkündür [38]. Bu nedenle, bankaların veya diğer finans kuruluşlarının, gelişen teknolojiyle birlikte artan imkanları kullanması donuk alacak olarak sınıflandırılması muhtemel kredilerin kullandırılmamasına yönelik olarak önemli ölçüde fayda sağlayacaktır.

3.3 BANKACILIK SEKTÖRÜNDE KARŞILAŞILABİLECEK RİSK TÜRLERİ

Risk, bir işlem ya da faaliyet nedeniyle parasal kaybın meydana gelmesi veya bir giderin ya da zararın oluşması durumunda ekonomik faydanın azalması olasılığını ifade etmektedir [39].

Bankacılık faaliyetleri kapsamında karşılaşılabilecek riskler genel olarak;

- kredi riski,
- likidite riski,

- operasyonel risk,
- faiz oranı riski,
- kur riski,
- ülke riski

olarak sıralanabilir.

Kredi Riski: Bir banka veya finansman kuruluşu ile anlaşmaya taraf olan kişinin sözleşme ile belirlenen koşullara uygun şekilde yükümlülüklerini yerine getirmemesi ihtimalidir. Kredi risk yönetiminin temel hedefi, uygun parametreler dahilinde karşılaşılabileceği riskleri yönetmek suretiyle bankanın risk ayarlı getirisini en üst seviyede tutmaktır [40].

Likidite Riski: Bankanın nakit ödemelerini gerçekleştirebilecek tutarda nakit mevcuduna sahip olmamasından kaynaklanan zarar ihtimalidir. Örneğin, nakit akışının öngörülememesi, kaynaklar ve yükümlülükler arasında vade uyumunun sağlanamaması nedeniyle fonlamada sıkıntı yaşanması.

Operasyonel Risk: Yetersiz iç süreçler, personel ve sistemlerden ya da harici olaylardan kaynaklanan ve yasal riski de kapsayan zarar etme olasılığıdır. Örnek olarak, banka veya müşteri varlıklarının zimmete geçirilmesi, çek sahtekarlıkları gösterilebilir.

Faiz Oranı Riski: Faiz oranlarında yaşanan değişimler nedeniyle bankanın alım-satım hesaplarında izlenen bir varlık veya finansal aracın değerinin azalması riskidir. Örnek olarak, faizlerin artış eğiliminde olması nedeniyle Bankanın menkul kıymet portföyünün değer kaybetmesi gösterilebilir.

Kur Riski: Döviz kurundaki değişimler neticesinde bankanın yabancı para üzerinden olan varlıkları ve borçları nedeniyle zarar etmesi ihtimalidir. Örneğin, kısa açık pozisyon taşıyan döviz kurunun diğer döviz kurları karşısında değer kazanması durumu.

Ülke Riski: Bankanın veya işlem yapılan diğer kişinin, farklı bir ülkedeki muhtelif değişikliklerden olumsuz yönde etkilenecek zarar etme ihtimali anlamına gelmektedir. Bir örnek olarak, başka bir ülkedeki ekonomik olumsuzluklar sebebiyle, ilgili ülkeyle ticaret yürüten borçlunun mali bünyesinin bozulması ve borcunu ifa edememesi gösterilebilir.

Genel olarak değerlendirildiğinde, veri madenciliği yöntemlerinin etkin kullanılması durumunda bankacılık sektöründe karşılaşılabilecek muhtemel risklere ilişkin öngörude bulunulabileceği görüşüne ulaşılmıştır.

3.4 KREDİ DEĞERLENDİRME SÜRECİNDE MEVCUT DURUM

1990'lı yılların sonuna kadar bankalara yapılan kredi başvuruları, kişilerin geçmiş ödeme performansları, gelir düzeyleri ve benzeri az sayıda kriteri dikkate alarak ve nispeten yüzeysel değerlendirmeler neticesinde sonuca ulaştırılmaktaydı. Bu nedenle, başvuru sahibinin talep ettiği krediyi ödeyip ödeyemeyeceğine dair detaylı bir öngörü elde edilememekteydi.

Güncel bankacılık uygulamalarında ise, kredi tahsisi aşamasında başta Kredi Kayıt Bürosu (KKB) tarafından temin edilen birçok veriden faydalanılmaktadır. 11 Nisan 1995 tarihinde dokuz bankanın ortaklığında kurulan Kredi Kayıt Bürosu'nun, gelinen aşamada bankalar, finansman şirketleri, leasing, faktöring, varlık yönetim şirketleri ve sigorta şirketleri olmak üzere yaklaşık 150 üyesi bulunmaktadır. 2011 yılında yapılan muhtelif yasal düzenlemeler neticesinde Türkiye Bankalar Birliği (TBB) nezdinde, kredi kuruluşları ile Bankacılık Düzenleme ve Denetleme Kurulu'nca uygun görülecek finansal kuruluşların müşterilerinin risk bilgilerini toplamak ve söz konusu bilgileri bu kuruluşlar ile gerçek veya tüzel kişilerin kendileriyle ya da onay vermeleri koşuluyla özel hukuk tüzel kişileri ve üçüncü gerçek kişiler ile de paylaşılmasını sağlamak üzere Risk Merkezi (RM) kurulmuştur. T.C.M.B. bünyesindeki Risk Santralizasyon Merkezi'nin devri neticesinde T.B.B. Risk Merkezi 2013 yılının Haziran ayında faaliyete geçmiştir. 2014 yılının Eylül ayı itibarıyla KKB, Findeks platformu ile bireylere ve reel sektöre yönelik çek raporu, risk raporu vb. hizmetleri tek bir çatı altında toplamıştır [41].

Findeks Kredi Notu; kişilerin kredi ödeme performanslarına, limit ve borç durumlarına, yeni kredili ürün açılışlarının yoğunluğu gibi detaylara göre hesaplanan ve kişinin gelecek 12 aydaki borçlarının ödenebilmesi olasılığını gösteren bir skordur. Findeks Kredi Notu 1-1900 arasında değişir. Söz konusu notun yüksek olması, risk seviyesinin düşük olduğu anlamına gelmektedir. Findeks Kredi Notu aşağıdaki unsurların ağırlıklandırılması ile oluşturulmaktadır;

- Kredi Ürün Ödeme Alışkanlıkları (%45): Kredi geri ödemelerinin sözleşmede belirlenen vadede gerçekleştirilmesi kredi notunu yükseltir.
- Mevcut Hesap ve Borç Durumu (%32): Kapanan kredilerin ödeme performansları, borçlarının teminat yapıları ve limitleri göz önünde bulundurularak hesaplamaya dahil edilmektedir.

- Yeni Kredili Ürün Açılışları (%5): Henüz ödeme performansı oluşmamasına karşın yakın tarihte kullanılan krediler risklilik oranını yükselten bir unsurdur.
- Kredi Kullanım Yoğunluğu (%18): Borcunu zamanında ifa ederek düzenli bir ödeme geçmişine sahip olan kişinin kredi notu, kredi kullanmayan veya az kredi kullanan kişiye göre daha yüksek olabilir [42].

BDDK tarafından yayımlanan iyi uygulama rehberlerinde bankaların;

- bireysel kredi tahsisi öncesinde;
 - kredinin kullanım amacı,
 - çalışma durumu ve gelir düzeyi,
 - geri ödeme potansiyelinin kaynağı,
 - ailesinin ve kendisine ekonomik olarak bağımlı kişilerin durumu,
 - finansal yükümlülükler ve bu yükümlülüklerin karşılanmasına yönelik masraflar,
 - düzenli harcamalar,
 - teminat,
 - varsa garanti gibi diğer risk azaltıcılar
- ticari kredi tahsisi öncesinde;
 - kredinin amacı,
 - gelir ve nakit akışı,
 - finansal pozisyonlar ve taahhütler,
 - iş modeli ve kurumsal yapı,
 - finansal tahminlerle desteklenmiş iş planları,
 - teminat,
 - varsa garanti gibi diğer risk azaltıcılar,
 - kredi türüne özgü yasal evrak

hakkında mümkün olduğu ölçüde uygun kanıtlarla desteklenen bilgilere sahip olması ve bu bilgileri kullanması gerektiği hususuna yer verilmiştir [43].

Gelinen aşamada bankalar tarafından kişi bazında çok sayıda veriden faydalanılıyor olmasına karşın, kredi tahsisi sürecinde benzer profildeki müşterilerin kredi geri ödeme performanslarının da dikkate alınmasının kredi kullandırılması öncesinde faydalı bir parametre olabileceği düşünülmektedir. Bilhassa veri madenciliği ve makine öğrenmesi uygulamalarının yaygınlaşması sayesinde günümüzde ödeme alışkanlıkları, borçluluk düzeyi, yeni kredili ürün açılışları, kredi

kullanım yoğunluđu gibi faktörlerin yanı sıra, başvuru sahibine benzer gelir/gider örüntüsüne sahip kişilerin kredi geri ödeme alışkanlıklarının değerdendirilmesi, başvuruda bulunan kişinin geri ödemeyi aksatma sonucunu doğurabilecek harcama alışkanlıkları ve benzeri unsurları dikkate alarak kredi kullandırma kararı alacak kişilere çok boyutlu analiz yapılması fırsatı sunabilmektedir.



BÖLÜM IV

BANKACILIKTA VERİ MADENCİLİĞİ İLE MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN KULLANIM ALANLARI VE BİR UYGULAMA

4.1 BANKACILIKTA VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN KULLANIMI

Finansal İstikrar Kurulu, makine öğrenmesi ve yapay zeka uygulamalarının;

- finansal kuruluşların kredi portföy kalitesinin değerlendirilmesi, sigorta poliçelerinin fiyatlaması ve müşteri etkileşiminin otomatize edilmesi,
- kuruluşların sermayeleri ile ve hacimli işlem pozisyonlarının optimizasyonu,
- hedge fonların, brokerların/dealerların ve benzer kuruluşların daha yüksek getirilerin tespitinde ve takas işlemlerinin optimizasyonunda,
- gerek kamu gerek özel sektörün mevzuata uygunluk, gözetim, veri kalitesinin değerlendirilmesi ve dolandırıcılık işlemlerinin tespitinde

kullanılabileceğini belirtmiştir [44].

Benzer şekilde, Uluslararası Finans Enstitüsü tarafından 2019 yılında yapılan araştırmada, finansal kuruluşların;

- kredi riski modelleme ve yönetimi aşamalarında makine öğrenmesi tekniklerinin bilhassa 2018 yılında kayda değer oranda arttığı,
- 2018 yılına kadar ağırlıklı olarak perakende portföylerin kredi değerliliği ve kurumsal portföylerin kredi izleme süreçlerinde kullanılan makine öğrenmesi yöntemlerinin, 2019 yılında küçük ve orta ölçekli işletmelere yönelik işlemlerde kullanımının önemli oranda arttığı,
- makine öğrenmesi tekniklerinin işletmelere model başarılarının artması, veri eksikliklerin ve tutarsızlıklarının tespitinin yanı sıra, yeni risk alanlarının tespitinde katkı sağladığı

belirtilmiştir [45].

Küresel finansal krizler sonrasında bankalarda yürütülen risk yönetimi faaliyetinin önemi artmış olup, riskin nasıl belirlendiği, ölçüldüğü, raporlandığı ve

yönetildiği konuları üzerinde belirgin bir odaklanma gözlemlenmektedir. Bu kapsamda, makine öğrenmesinin iş uygulamaları alanında kullanımı kayda değer ölçüde etkilenmiştir. Yasal regülasyonların genişlemesi ve derinleşmesi, müşteri beklentilerinin değişmesi ve risk türlerinin değerlemesinin risk yönetimindeki değişimi yönlendirmesi beklenmektedir. Makine öğrenmesi, karmaşık ve lineer olmayan örüntü halindeki veri setleri üzerinden daha tutarlı risk modellemesi yapılabilmesine olanak sağlayabilir [46].

Bir diğer araştırmada yer verildiği üzere, bankacılık sektöründe veri madenciliği kullanılan alanlar;

- hile (yolsuzluk) tespiti,
- pazarlama faaliyetleri,
- risk yönetimi,
- müşteri ilişkileri yönetimi

olarak sıralanabilir [47].

İlaveten, veri madenciliğinin finans sektöründe görev yapan kişilerce kaynağın daha verimli kullanılmasına, sistemin kontrol edilebilmesine, riskin azaltılmasına veyahut transfer edilmesine, müşteri ihtiyaçlarının anlaşılabilmesine, etkin kaynak yönetimine, pazar payı artışına katkı sağladığını savunmakla birlikte, ayrıca, bilhassa bankacılık ve sigortacılık sektörlerinde temerrüt riskini ve risk gruplarını belirlemekte, her bir müşteri bazında optimal sigorta seçeneklerini sınıflandırmakta, müşteri memnuniyetini artırmakta ve kredi kartı suiistimallerini tespit etmekte kullanılabileceğini savunulmaktadır [48].

Genel olarak değerlendirildiğinde, gerek bankaların veri tabanında yer alan veri miktarında olağanın üzerinde yaşanan artış gerekse finansal piyasalarda karşılaşılan risk türlerinin ve söz konusu risklerin olası olumsuz etkilerinin artması nedenleriyle bankacılık sektöründe veri madenciliği yöntemlerinin uygulanmasının kaçınılmaz olduğu sonucuna ulaşılmıştır.

4.2 UYGULAMA

Çalışmamızda Türk Bankacılık Sektörüne yönelik detaylı bilgiler verilmiş olması nedeniyle, uygulama bölümünde Türkiye’de faaliyet gösteren bankalar tarafından kullanılan kredilere ilişkin bir veri seti üzerinden analiz yapılması öngörülmüştür. Öte yandan, 6698 numaralı Kişisel Verilerin Korunması Kanunu’nda yer alan hükümler nedeniyle bankalar kredi müşterilerine ilişkin herhangi bir veri

setini tarafımızla paylaşmamıştır. Bu sebeple çalışmamızın uygulama bölümünde, Amerika Birleşik Devletleri merkezli Lending Club isimli firma tarafından 2007-2020 yılları arasında kredi kullandırılan kişilere ait veriler “sınıflandırma” algoritmalarının performanslarının karşılaştırılması amacıyla kullanılmıştır.

Lending Club, fon ihtiyacı olan kişiler ile fon fazlası olan kişileri bir araya getiren ilk “benzerler(eşler) arası (peer-to-peer)” finansman kuruluşudur. Bu sistemde yatırımcılar Lending Club firmasının internet sitesinde yer alan kredi listelerini araştırarak borçlu, kredi miktarı, kredi notu ve kredi amacı hakkında edindiği bilgiler neticesinde yatırım yapmak istedikleri kredileri seçebilmektedir. Fon temini aşamasında banka vb. aracı kuruluşların bulunmaması nedeniyle, ihtiyacı olan kişiler nispeten düşük oranda faiz karşılığında fon sağlayabilmektedir. Bir örnek vermek gerekirse, teminatsız kredilerin yıllık %9 faiz oranı ile kullanıldığı bir piyasada Lending Club vasıtasıyla fon ihtiyacının karşılanması durumunda faiz oranı %6 veya %7 olarak belirlenebilmektedir. 2015 yılı sonu itibarıyla toplam 15,98 milyar dolar tutarında kredi kullandıran Lending Club, 2020 yılında Radius Bank’ı satın almıştır.

Tez çalışmamız kapsamında verinin düzenlenmesinde, algoritmaların çalıştırılmasında ve verilerin analizinde “Phyton” programı kullanılmıştır. Ayrıca, uygulamamızda çalışmamızın önceki başlıklarında detaylı olarak yer verilen CRISP-DM (Cross Industry Standart Process for Data Mining) metodu kullanılmıştır. Bu metoda göre izlenen aşamalar;

- I. Aşama: Çalışmanın Tanımlanması (Business Understanding)
- II. Aşama: Verinin Tanımlanması (Data Understanding)
- III. Aşama: Verinin Hazırlanması (Data Preparation)
- IV. Aşama: Modelleme (Modelling)
- V. Aşama: Değerlendirme (Evaluation)
- VI. Aşama: Dağıtım (Deployment)

başlıkları halinde aşağıda detaylı olarak açıklanmıştır.

4.2.1 Çalışmanın Tanımlanması

Bankaların/finans kuruluşlarının sermaye yapılarını kuvvetli tutabilmeleri amacıyla aktif kalitelerini iyileştirmeleri ve bu sebeple kullandırılan kredilerin temerrüt oranını mümkün olduğu kadar düşük tutmaları gerekmektedir. 1990’lı yılların sonuna kadar gelir düzeyi, kredi ödeme performansı ve benzeri birkaç veriden faydalanılarak sonuçlandırılan kredi tahsis süreçleri günümüzde, veri madenciliği ve

makine öğrenmesi yöntemlerinin kullanımının artmasıyla, daha sağlıklı analiz yapabilmekte ve kredinin temerrüt olasılığı hakkında daha başarılı öngörülerde bulunulmasına olanak sağlamaktadır.

Çalışmamız kapsamında, makine öğrenmesi tekniklerinin kredi tahsisi sürecindeki temerrüt olasılığının tahmin performanslarının karşılaştırılması amaçlanmıştır.

Çalışma sonucunda elde edilecek veriler, makine öğrenmesi algoritmalarının performanslarının değerlendirmesinde yaygın olarak kullanılan aşağıdaki kriterler çerçevesinde değerlendirilecektir.

i. Karmaşıklık Matrisi (Confusion Matrix):

Sınıflandırma modellerinin performansının değerlendirilmesi amacıyla kullanılan karmaşıklık matrisi, sınıflandırma işlemi sonucunda ulaşılan tahmini veriler ile gerçek sınıf etiketleri arasındaki ilişkiyi göstermektedir. Karmaşıklık matrisinde yer alan değerler aşağıdaki şekilde hesaplanmaktadır.

- True Positive (TP): Gerçek sınıfı pozitif olarak tahmin edilen örnek sayısı.
- True Negative (TN): Gerçek sınıfı negatif olarak tahmin edilen örnek sayısı.
- False Positive (FP): Gerçek sınıfı negatif olmasına karşın pozitif olarak tahmin edilen örnek sayısı.
- False Negative (FN): Gerçek sınıf pozitif olmasına karşın negatif olarak tahmin edilen örnek sayısı.

Model performansının anlaşılabilmesi amacıyla kullanılan aşağıdaki metrikler, karmaşıklık matrisinde yer verilen değerler aracılığıyla hesaplanmaktadır.

- Doğruluk (Accuracy): $(TP+TN)/(TP+TN+FP+FN)$
- Hassasiyet (Precision): $TP/(TP+FP)$
- Duyarlılık (Recall): $TP/(TP+FN)$

Karmaşıklık matrisi ve yukarıda belirtilen metrikler, modelin hangi sınıfları ne ölçüde doğru tahmin ettiğini analiz edebilmek için kullanılmaktadır.

ii. ROC Eğrisi:

Sınıflandırma problemlerinde ve bilhassa makine öğrenimi modellerinin performansını değerlendirmede kullanılan bir araç olan ROC eğrisine (Receiver Operating Characteristic), genellikle ikili sınıflandırma problemlerinde başvurulmaktadır.

ROC eğrisi, X ekseninde yanlış pozitif oranının [False Positive Rate (FPR=FP/(FP+TN))] Y ekseninde ise gerçek pozitif oranının [True Positive Rate (TPR=TP/(TP+FN))] bulunduğu grafikdir. Eğrinin altındaki alan arttıkça sınıflar arasında ayırt etme performansı yükselmektedir. Sol üst köşede bulunan (0 FPR, 1 TPR) noktasına en yakın olan eğri ideal ROC eğrisidir. Bu noktada TPR (True Positive Rate) gerçek pozitif oranıdır ve duyarlılığı temsil eder.

iii. AUC Skoru:

ROC eğrisi altında kalan alanı ifade eden AUC skoru (Area Under the Curve), modelin genel performansını ölçen bir metriktir. 0 ile 1 arasında değer alabilen AUC skoru 1 değerine yaklaştıkça sınıflandırma performansının iyileştiği sonucuna ulaşılır.

4.2.2 Verinin Tanımlanması

Google Inc. bünyesinde faaliyet gösteren bir veri bilimi platformu olan “kaggle.com” isimli veri sağlayıcısından temin edilen veri seti, Lending Club firmasının 2007-2020 yılları arasında kullandığı yaklaşık 2.925.493 adet bireysel ve ticari nitelikli krediye ilişkin bilgiler içermektedir. Eksik veri içeren kayıtların temizlenmesi sonucunda kredi sayısı 1.363.544 olarak şekillenmiştir.

Çalışmamız kapsamında, kredi kullanan kişilerin;

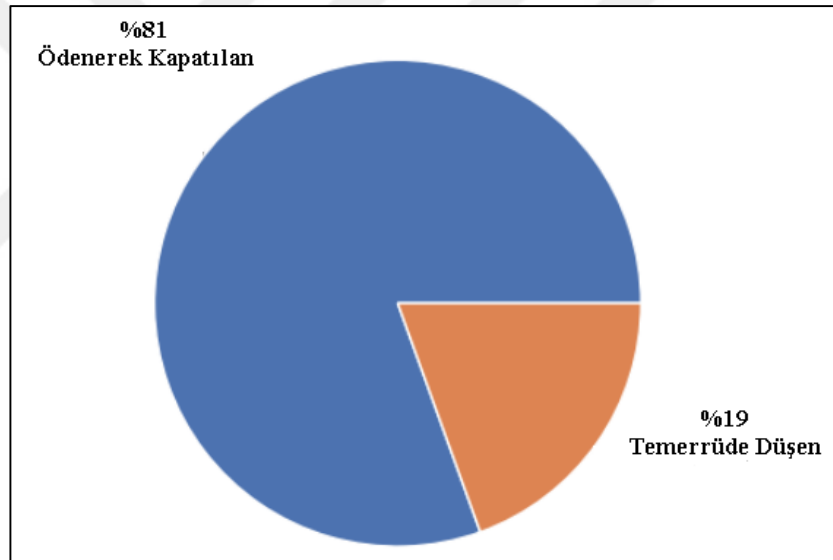
- beyan edilen yıllık gelir düzeyi,
- son bir yıl içerisinde temerrüde düşen kredi sayısı,
- son iki yıl içerisinde ödeme yükümlülüklerini 30 gün veyahut daha fazla sürede gecikmeli olarak yerine getirdikleri kredi sayısı,
- aylık ödenmesi gereken kredi tutarının, aylık fiilen ödenen kredi tutara oranı,
- en eski kredi kullanım tarihi,
- kredi kullandıkları tarih boyunca şekillenen en yüksek ve en düşük kredi skorları,
- kredi notu (borçlunun kredi geçmişine, teminatın kalitesine, anapara ve faizin geri ödenme olasılığına dayalı olarak bir kredi için belirlenen puan),
- ikamet edilen konutun sahipliği,
- kredinin kullanılması durumunda borçlunun aylık toplam ödeme tutarı,
- kredi faiz oranı,
- kredi tutarı,

- kredinin ödenme durumu (kredinin ödenerek kapatılması veya temerrüde düşmesi),
- açık kredi sayısı,
- iflas kayıtları,
- toplam rotatif kredi bakiyesi,
- kredi vadesi

dikkate alınarak analiz tamamlanmıştır. Analizde kullanılan bazı parametrelere ilişkin detaylı açıklamalara aşağıda yer verilmiştir.

- Ödenerek Kapatılan/Temerrüde Düşen Kredi Oranı

Veri setinde yer alan kredilerden sözleşmede öngörülen şekilde ödenerek kapatılan veyahut ödenmeyerek temerrüde düşen kredi oranlarına aşağıdaki şekilde yer verilmiştir.

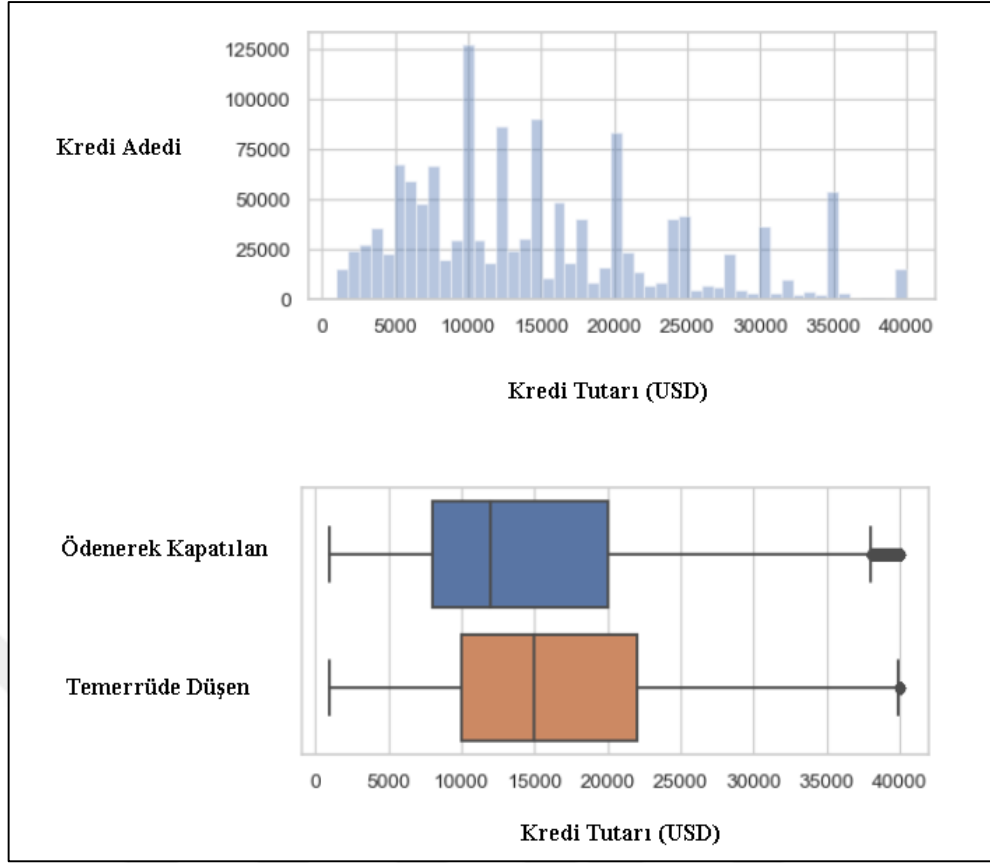


Şekil 8: Ödenme Durumuna Göre Kredilerin Dağılımı

Yukarıdaki grafikten görüleceği üzere, veri setinde yer alan kayıtların %81'i ödenerek kapatılmış olup, kalan bölüm ise temerrüde düşmüştür.

- Kredi Tutarına Yönelik Açıklamalar

Kredi tutarına ilişkin bilgiler aşağıdaki şekillerde yer almaktadır.



Şekil 9: Kredi Tutarına İlişkin Grafikler

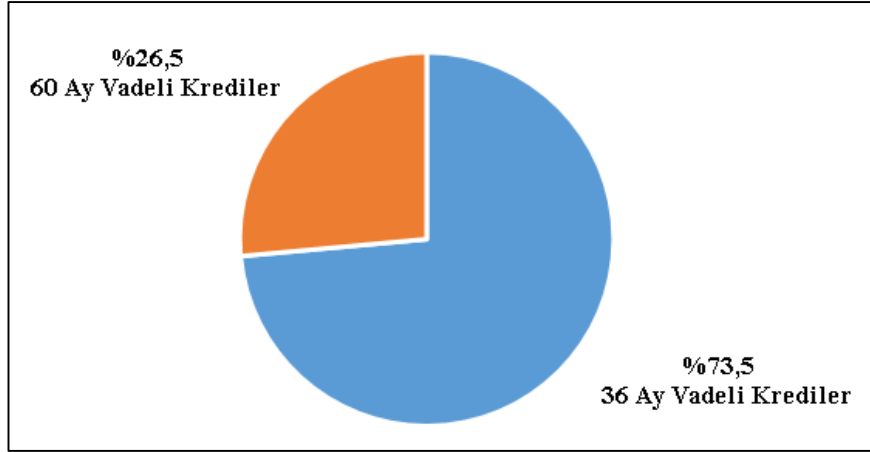
Grafiklerden görüleceği üzere kullanılan veri setinde;

- kredi miktarının 10.000 USD limitli kredilerde yoğunluk gösterdiği, diğer taraftan 40.000 USD limite kadar krediler kullandırıldığı,
- daha yüksek limitli kredilerin temerrüde düşme oranının nispeten yüksek olduğu

anlaşılmaktadır.

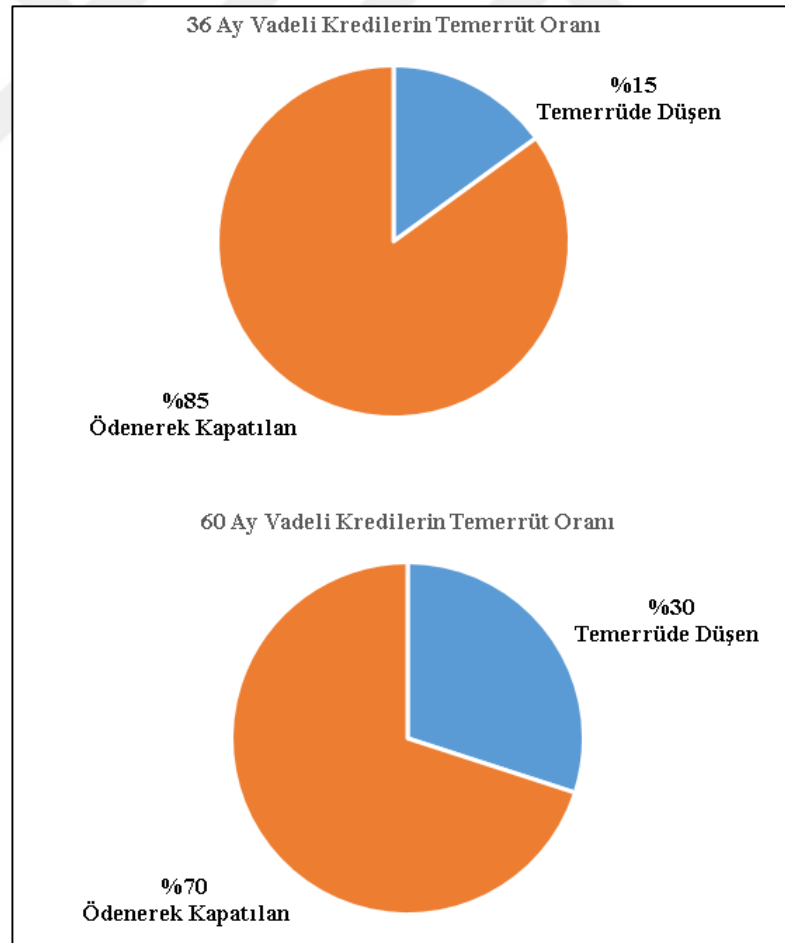
- Kredi Vadelerine Yönelik Açıklamalar

Veri setinde yer alan kredilerin vadelerine göre ödenme durumlarına ilişkin bilgilere aşağıdaki şekillerde yer verilmiştir.



Şekil 10: Vadelerine Göre Kredilerin Dağılımı

Şekilden anlaşılacağı üzere, veri setinde yer alan kredilerin önemli bölümü 36 ay vadeli kredilerden oluşmaktadır. Vadelerine göre kredilerin temerrüde düşme ve ödenerek kapatılma oranlarına ise aşağıdaki grafiklerde yer verilmiştir.

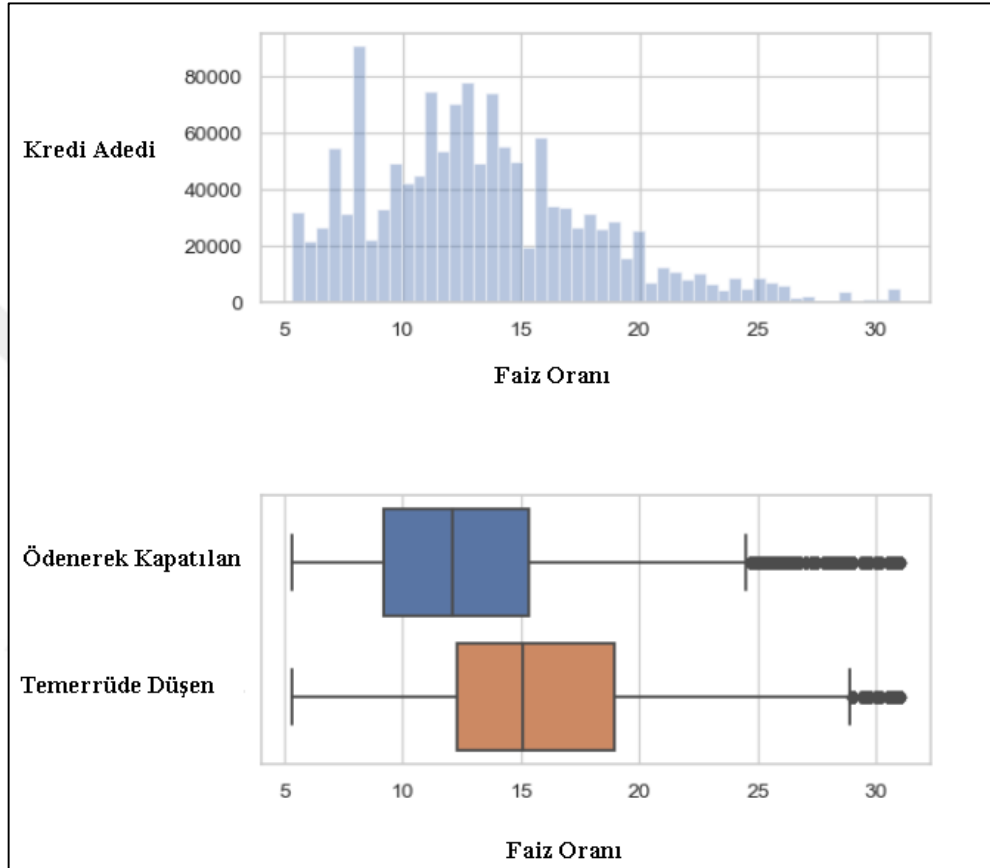


Şekil 11: Vadelerine Göre Kredilerin Temerrüt Oranı

Grafiklerden de anlaşılacağı üzere, 60 ay vadeli kredilerin temerrüt oranı 36 ay vadeli kredilerden daha yüksek şekillenmiştir.

- Faiz Oranlarına İlişkin Açıklamalar

Veri setinde yer alan kredilerin faiz oranlarına ilişkin bilgilere aşağıda yer verilmiştir.

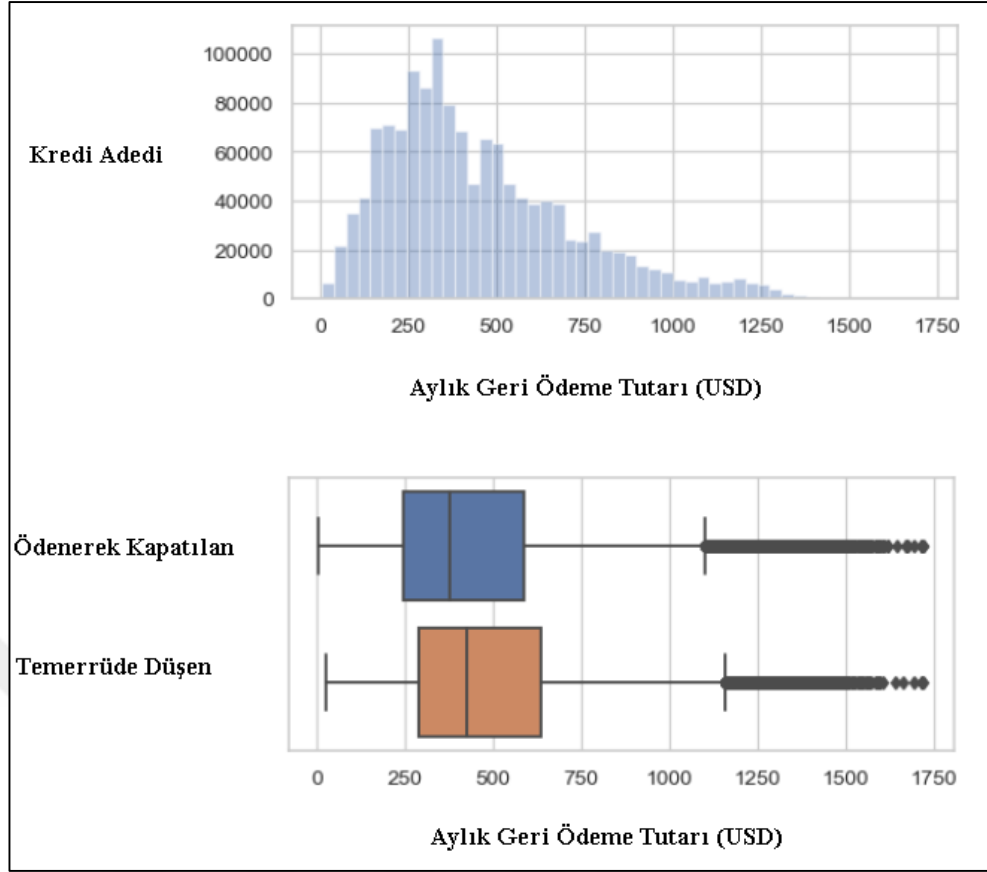


Şekil 12: Faiz Oranına İlişkin Grafikler

Şekillerden anlaşılacağı üzere, veri setinde yer alan kredilerden faiz oranı yüksek olanların temerrüde düşme oranının daha yüksek olduğu sonucuna ulaşılmaktadır.

- Kredilerin Aylık Geri Ödeme Tutarlarına Yönelik Açıklamalar

Veri setinde yer alan kredilerin aylık geri ödeme miktarlarına yönelik analiz bilgilerine aşağıdaki grafikte yer verilmiştir.

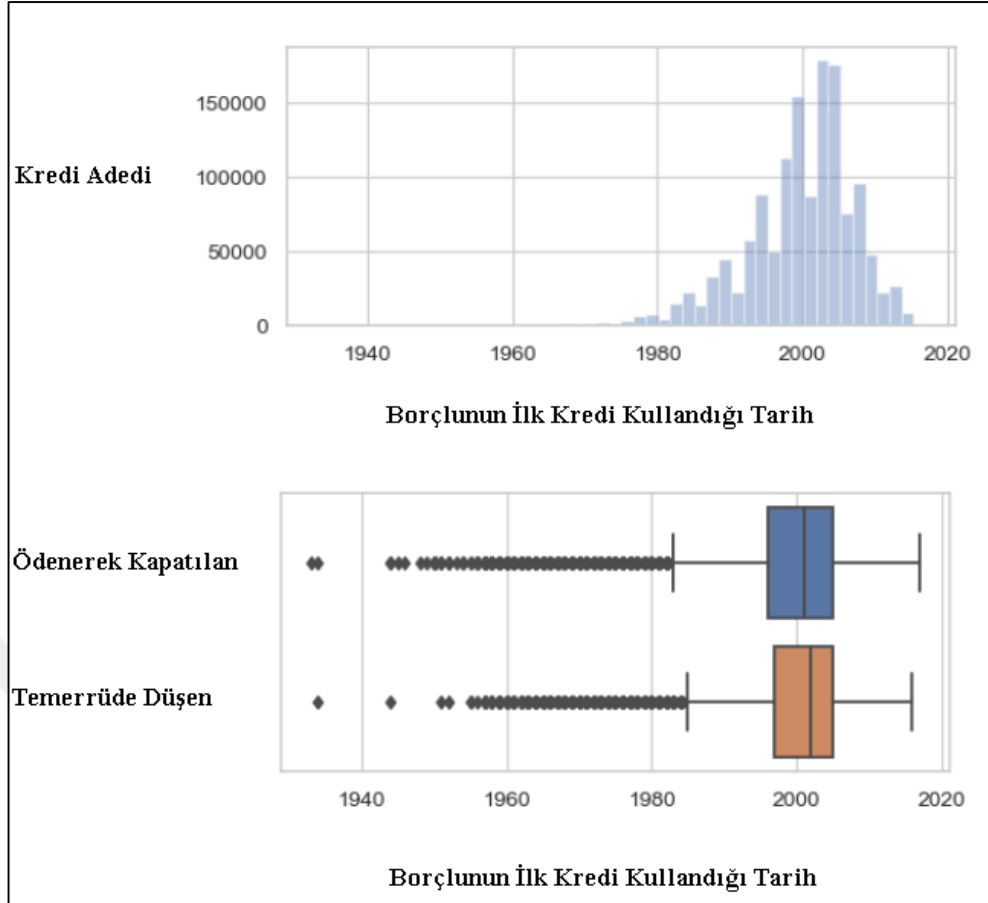


Şekil 13: Aylık Geri Ödeme Tutarına İlişkin Grafikler

Yukarıda yer alan şeklin yorumlanması neticesinde, ödenerek kapatılan kredilerin aylık geri ödeme tutarının, temerrüde düşen kredilerin aylık geri ödeme tutarından düşük olduğu anlaşılmaktadır.

- Borçluların İlk Kredi Kullandığı Tarihlere İlişkin Açıklamalar

Aşağıdaki grafiklerde borçluların ilk kredi kullandıkları tarih ile temerrüde düşme bilgilerine yer verilmiştir.

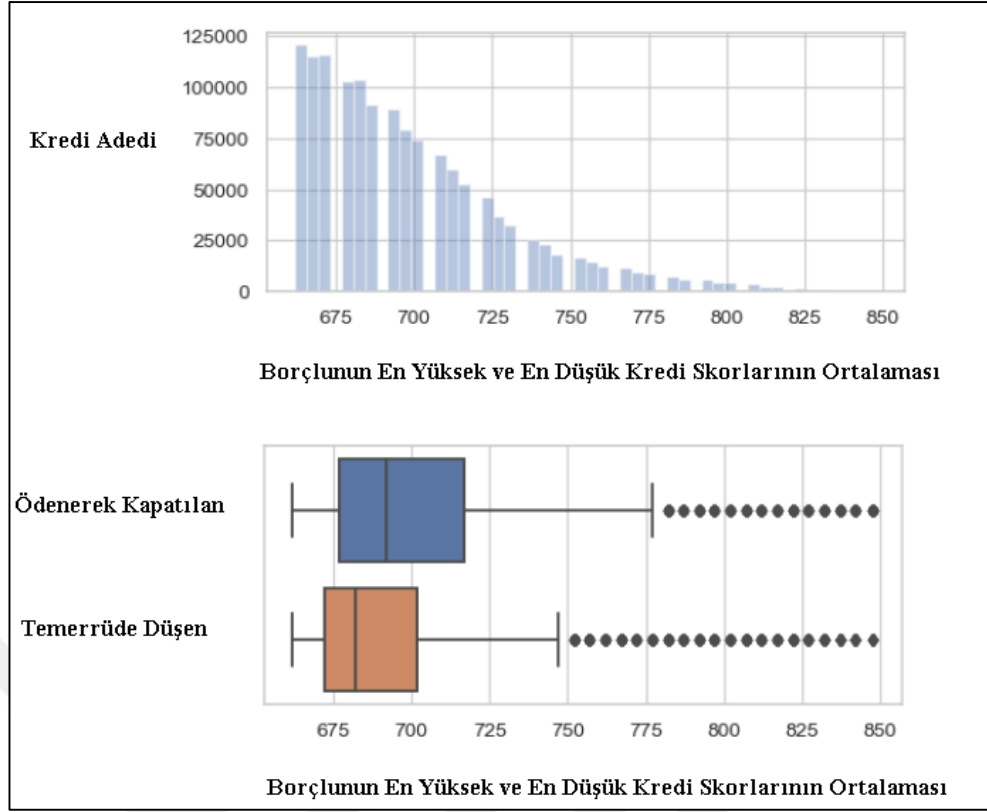


Şekil 14: Borçluların İlk Kredi Kullandığı Tarihlerle İlişkin Grafikler

Grafikten anlaşılacağı üzere, ilk kredi kullandıkları tarih eski olan borçluların kullandıkları yeni krediyi ödeyerek kapatma oranının nispeten yüksek olduğu belirlenmiştir.

- Kredi Başvuru Puanına Yönelik Açıklamalar

Kredi tahsisi aşamasında önemli bir veri olarak dikkate alınan kredi başvuru skorları ile kredilerin temerrüde düşmelerine veya ödenerek kapatılmalarına ilişkin verilere aşağıdaki grafiklerde yer verilmiştir. Değerlenen grafiklerde borçluların en yüksek ve en düşük kredi notlarının (FICO Score) ortalamaları dikkate alınmıştır.



Şekil 15: Borçlu Kredi Skorlarına İlişkin Grafikler

Grafikten anlaşılacağı üzere, kredi skoru yüksek olan borçlulara kullandırılan kredilerin ödenerek kapanma oranı daha yüksek şekillenmiştir.

4.2.3 Verinin Hazırlanması

Veri madenciliği sürecinin kritik öneme sahip olan aşamalarından bir tanesi, temin edilen verinin konsolidasyonu, temizlenmesi ve bulunması durumunda sapma içeren verilerin ayıklanması neticesinde verinin üzerinde analiz yapılabilecek bir aşamaya getirilmesidir.

Çalışmamız kapsamında kullanılan ham veri yaklaşık 2.925.493 adet krediye ilişkin bilgiler içermekte olup, eksik veri içeren kayıtların temizlenmesi sonucunda kredi sayısı 1.363.544 olarak şekillenmiştir. Öte yandan, söz konusu kredilerin kaç müşteriye ait olduğuna dair bir bilgi veri setinde yer almamaktadır.

4.2.4 Modelleme

Modelleme aşamasında çalışmamızın önceki bölümünde detaylarına yer verilen LightGBM, Lojistik Regresyon, K-En Yakın Komşu, Karar Ağaçları ve Yapay Sinir Ağları algoritmaları kullanılmıştır. Çalışmamız kapsamında veri setinin %80'lik

bölümü öğrenme (train), %20'luk kısmı ise test amaçlı kullanılmış olup, söz konusu algoritmalar rastgele seçilen örnekler üzerinden 20 defa çalıştırılmış ve aşağıdaki sonuçlara ulaşılmıştır.

Tablo 2: Uygulamada Kullanılan Algoritma Sonuçları

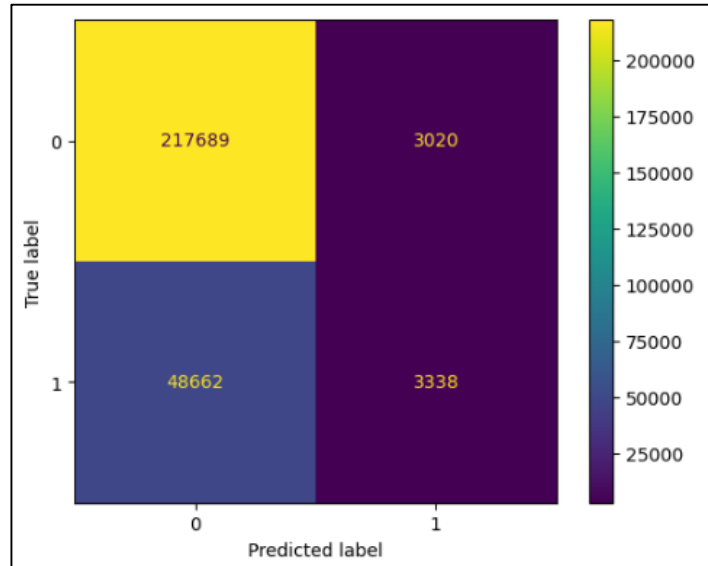
Kullanılan Algoritma	Doğruluk (%)	Hassasiyet (%)	Duyarlılık (%)
LightGBM	81,04	0,98	0,81
Lojistik Regresyon	80,9	0,99	0,81
Yapay Sinir Ağları	80,9	0,99	0,8
K-NN	78,1	0,94	0,81
Karar Ağaçları	78,06	0,92	0,82

Tablodan görüleceği üzere, doğruluk oranında LightGBM, hassasiyet oranında ise lojistik regresyon ve yapay sinir ağları, duyarlılık oranında ise karar ağacı algoritmaları düşük farklarla da olsa ön plana çıkmaktadır.

Çalışmamız kapsamında algoritmalar tarafından elde edilen sonuçların detaylarına başlıklar halinde aşağıda yer verilmiştir.

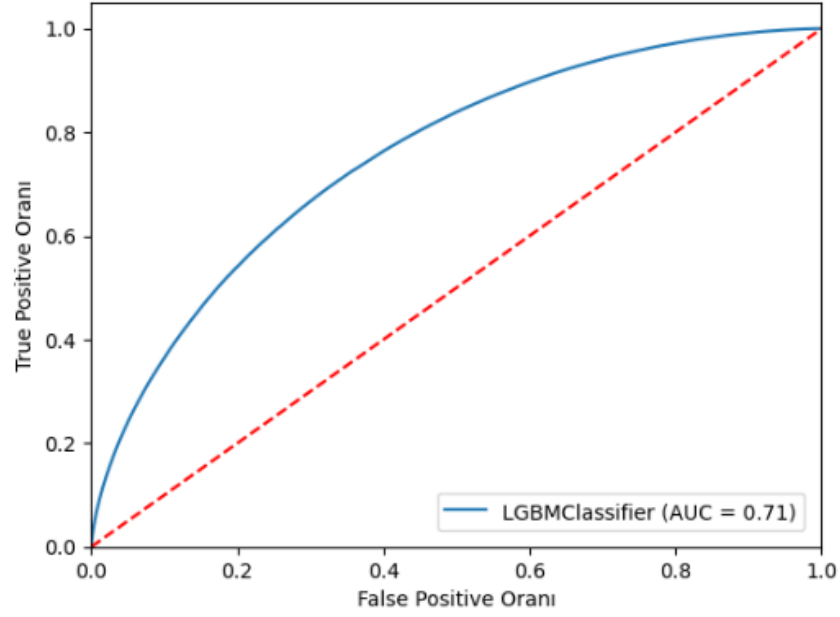
4.2.4.1 LightGBM

LightGBM algoritması kullanılarak oluşturulan karmaşıklık matrisi aşağıda yer almaktadır.



Şekil 16: LightGBM Karmaşıklık Matrisi

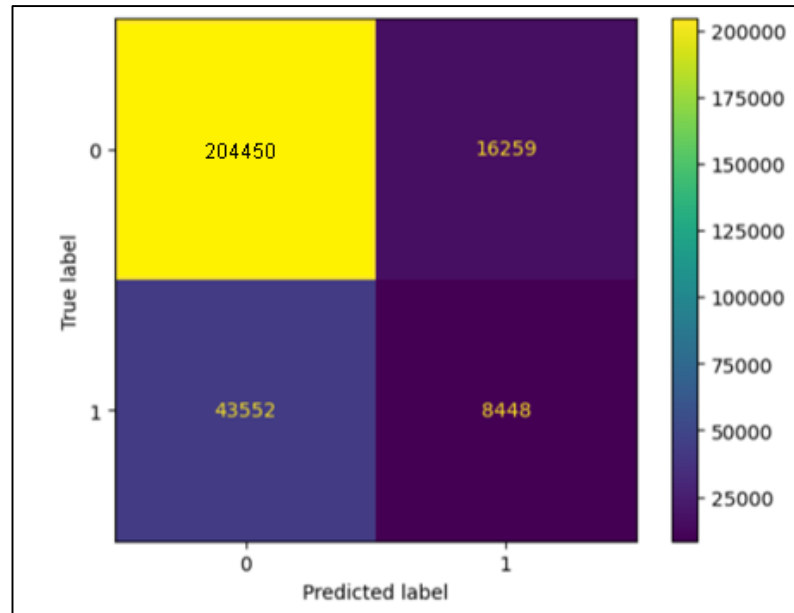
AUC skoru 0,71 olarak şekillenen modelimize konu ROC eğrisi aşağıdaki şekilde çizilecektir.



Şekil 17: LightGBM ROC Eğrisi

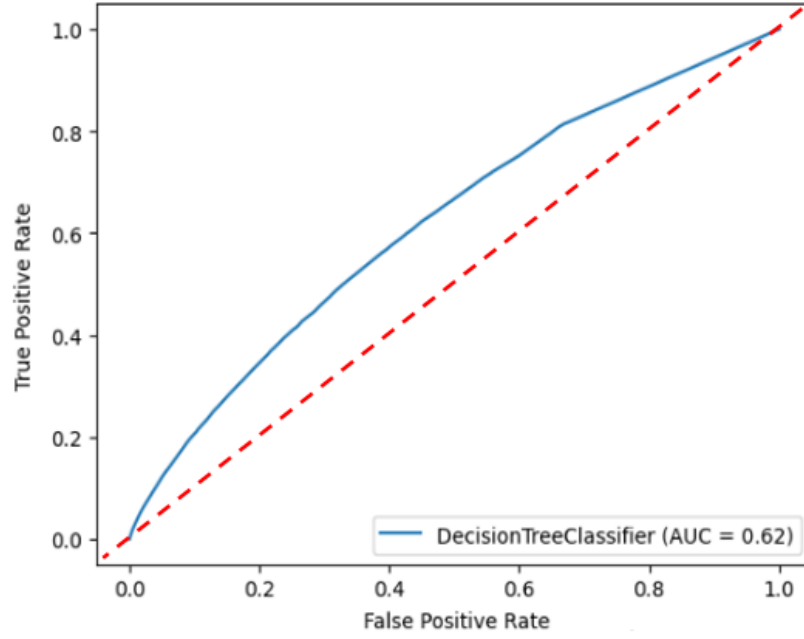
4.2.4.2 Karar Ağacı

Karar ağacı algoritması kullanılarak elde edilen karmaşıklık matrisine aşağıda yer verilmiştir.



Şekil 18: Karar Ağacı Karmaşıklık Matrisi

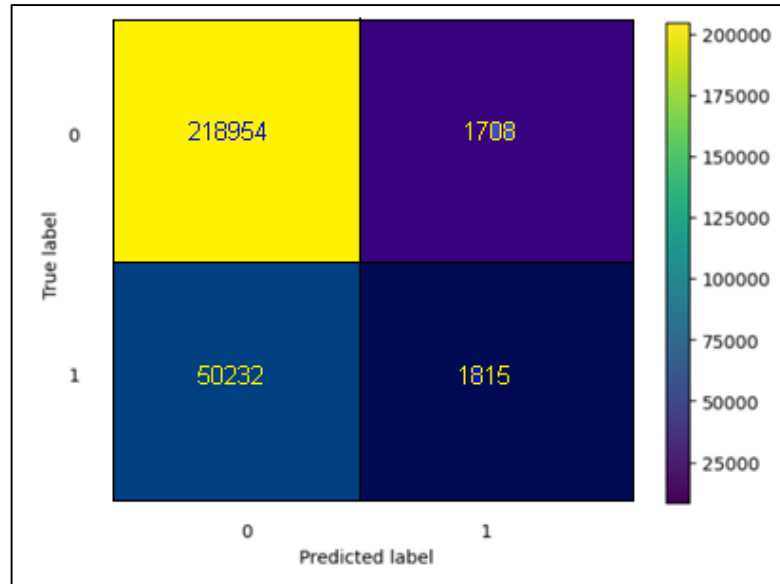
AUC skoru 0,62 olarak şekillenen modelimize konu ROC eğrisi aşağıdaki şekilde çizilecektir.



Şekil 19: Karar Ağacı ROC Eğrisi

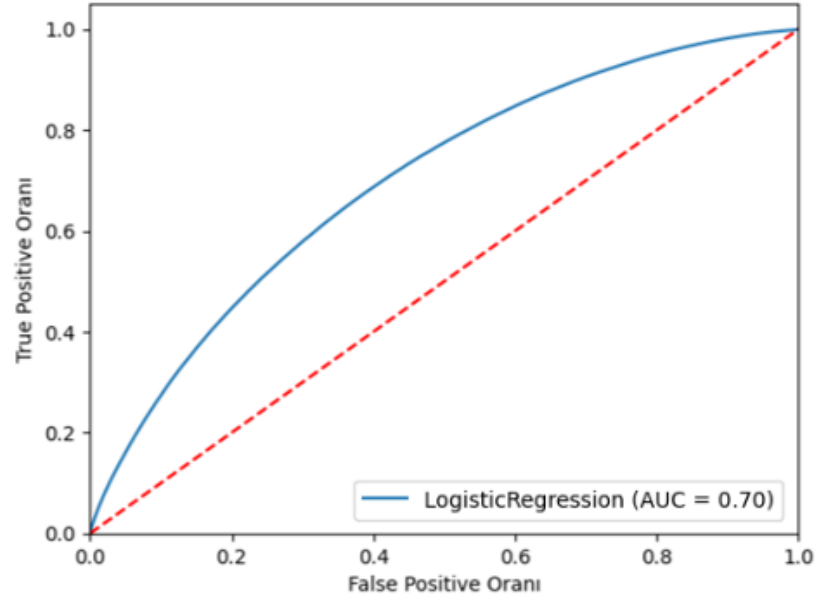
4.2.4.3 Lojistik Regresyon

Lojistik regresyon algoritması kullanılarak oluşturulan karmaşıklık matrisi aşağıda yer almaktadır.



Şekil 20: Lojistik Regresyon Karmaşıklık Matrisi

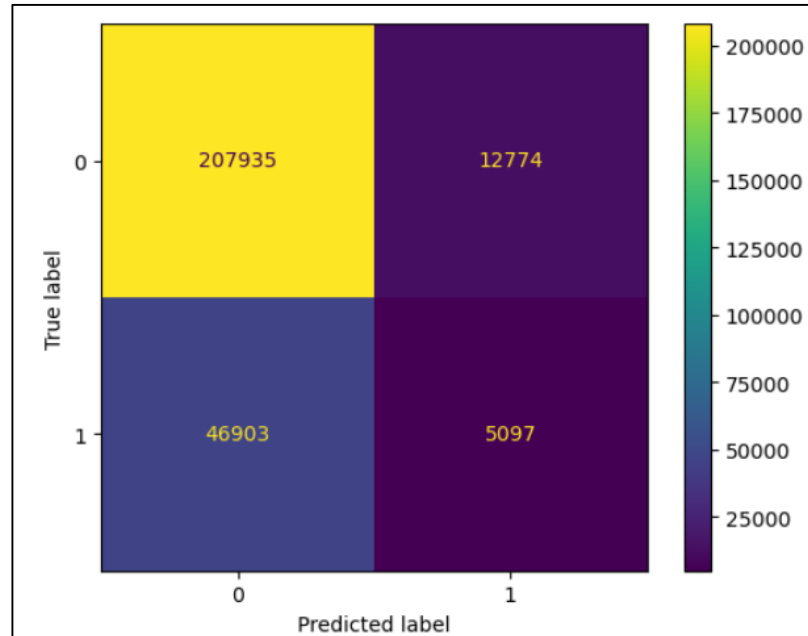
AUC skoru 0,7 olarak şekillenen modelimize konu ROC eğrisi aşağıdaki şekilde çizilecektir.



Şekil 21: Lojistik Regresyon ROC Eğrisi

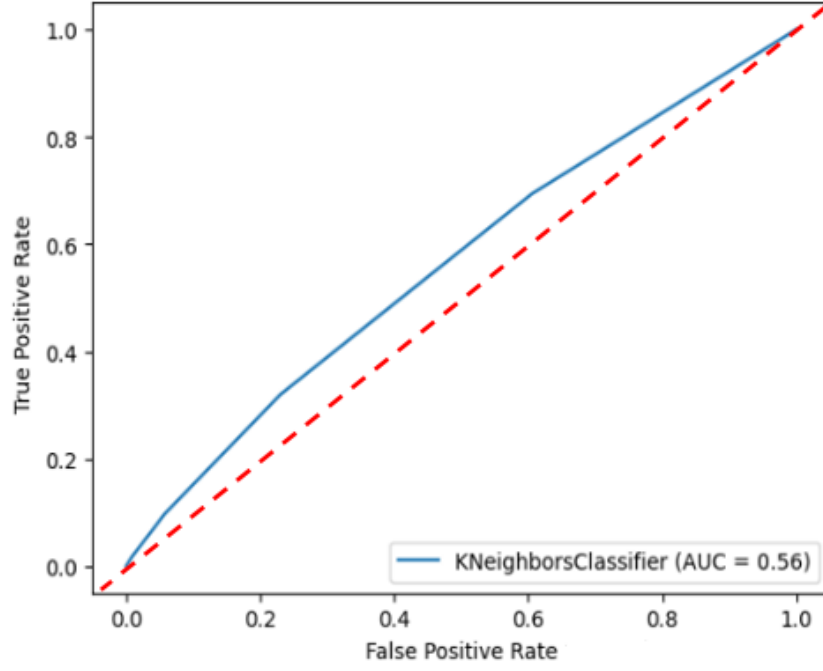
4.2.4.4 K-En Yakın Komşu

K-En Yakın Komşu algoritması kullanılarak oluşturulan karmaşıklık matrisi aşağıda yer almaktadır.



Şekil 22: K-En Yakın Komşu Karmaşıklık Matrisi

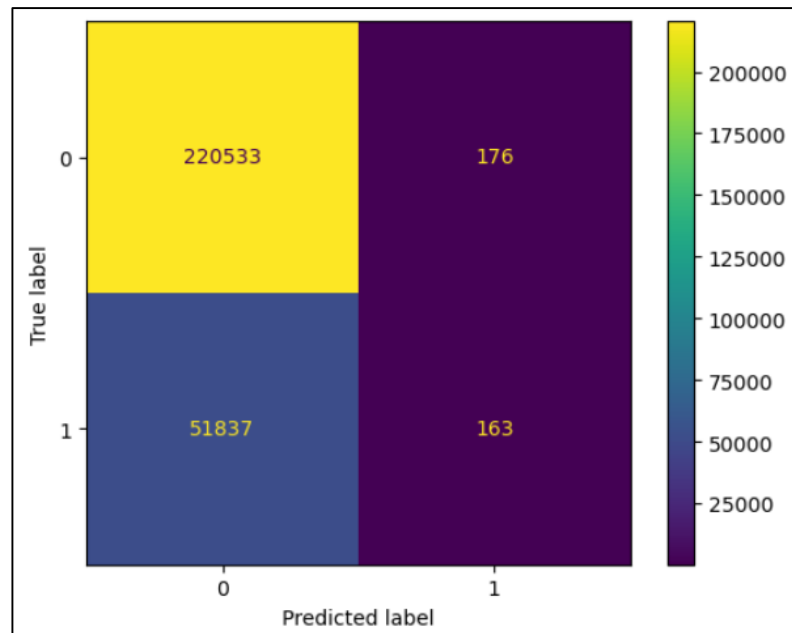
AUC skoru 0,56 olarak şekillenen modelimize konu ROC eğrisi aşağıdaki şekilde çizilecektir.



Şekil 23: K-En Yakın Komşu ROC Eğrisi

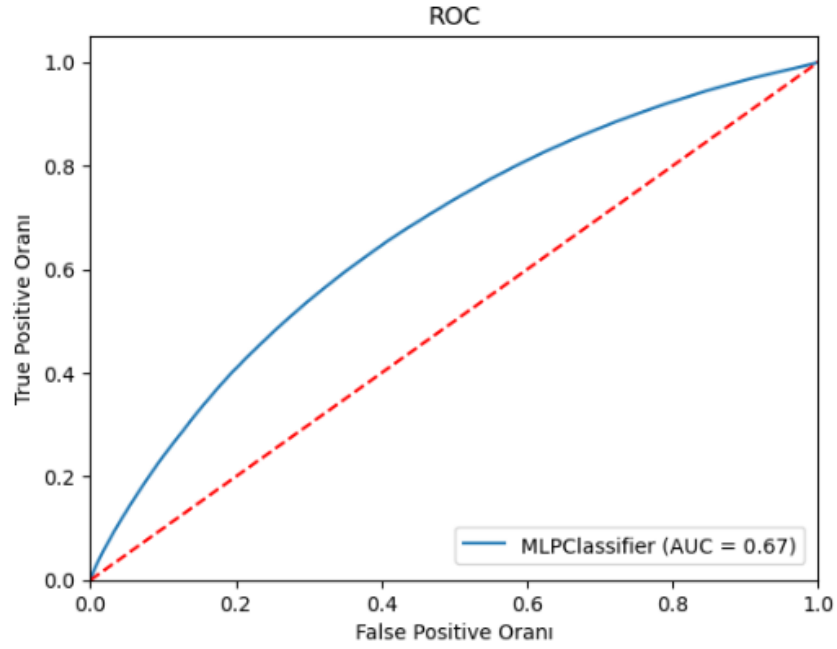
4.2.4.5 Yapay Sinir Ağları

Yapay sinir ağları algoritması kullanılarak oluşturulan karmaşıklık matrisi aşağıda yer almaktadır.



Şekil 24: Yapay Sinir Ağları Karmaşıklık Matrisi

AUC skoru 0,67 olarak şekillenen modelimize konu ROC eğrisi aşağıdaki şekilde çizilecektir.



Şekil 25: Yapay Sinir Ağları ROC Eğrisi

4.2.5 Değerlendirme

Genel olarak değerlendirildiğinde çalışmamız kapsamında kullanılan algoritmaların birbirine yakın sonuçlar elde ettiği, öte yandan LightGBM algoritmasının %81,04 doğruluk oranı ile öne çıktığı sonucuna ulaşılmıştır.

Karmaşıklık matrisinin yorumlanması neticesinde LightGBM algoritması tarafından;

- temerrüde düşmeyeceği öngörülen kredilerden 217.689 adedinin gerçekte temerrüde düşmediği (True Positive),
- temerrüde düşeceği öngörülen 3.338 kredinin temerrüde düştüğü (True Positive),
- temerrüte düşmeyeceği öngörülen 48.662 kredinin temerrüde düştüğü (False Negative),
- temerrüte düşeceği öngörülen 3.020 kredinin temerrüde düşmediği (False Positive)

anlaşılmaktadır. Dolayısıyla, LightGBM algoritmasının test amaçlı kullanılan 272.709 adet kredinin 221.027 adedinin temerrüde düşeceğini veya düşmeyeceğini doğru tahmin ettiği, kalan 51.682 krediye ilişkin tahminin ise hatalı olduğu tespit edilmiştir.

4.2.6 Sonuç ve Tartışma

Veri miktarının ve veriye ulaşma imkanlarının her geçen gün arttığı günümüz dünyasında, ihtiyaç duyulan verinin doğru olarak temini ve söz konusu veri kullanılarak analiz yapılabilmesi hususları önem kazanmaktadır.

Bankalar, müşteri sayıları ve veri miktarı hızlı artan işletmelerin başında gelmektedir. Rekabetçi piyasalarda faaliyet gösteren bankaların, pazarlama, kredi tahsisi, personel ve müşteri analizi vb. birçok alanda kullandıkları geleneksel yöntemlerin yetersiz kalması kuvvetle muhtemeldir. Bu nedenle teknolojik olanaklardan azami ölçüde faydalanması gereken bankaların, günümüzde veri madenciliği ve makine öğrenmesi metotlarına başvurmasının kaçınılmaz olduğu düşünülmektedir. Çalışmamız kapsamında, bankaların makine öğrenmesi yöntemlerini kullanmaları durumunda kredi tahsis süreçlerinde elde edecekleri fayda araştırma konusu yapılmıştır.

Çalışmamızda Türk Bankacılık Sektörüne ilişkin veriler ve yasal düzenlemeler aktarılmış olmasına karşın, ülkemizde faaliyet gösteren bankaların 6698 numaralı Kişisel Verilerin Korunması Kanunu hükümlerini gözeterek veri paylaşımı yapmamaları nedeniyle, Amerika Birleşik Devletleri menşeli Lending Club isimli finansman kuruluşuna ait veriler kullanılarak analiz yapılmıştır. Söz konusu finansman kuruluşu tarafından 2007-2020 yılları arasında kullandırılan yaklaşık 2.925.493 adet krediye ilişkin bilgiler üzerinden yapılan çalışmamızda eksik veri içeren kayıtlar temizlenmiş ve nihai olarak 1.363.544 adet krediye ilişkin bilgiler üzerinden analiz tamamlanmıştır. Veriye ilişkin işlemlerin ve analizlerin tamamı “Phyton” programlama dili kullanılarak yürütülmüştür. Ayrıca çalışmamızda, veri setinde yer alan kredileri kullanan kişilerin; beyan edilen yıllık gelir düzeyi, son bir yıl içerisinde temerrüde düşen kredi sayısı, son iki yıl içerisinde ödeme yükümlülüklerini 30 gün veya daha fazla sürede gecikmeli olarak yerine getirdikleri kredi sayısı, aylık ödenmesi gereken kredi tutarının aylık fiilen ödenen kredi tutarına oranı, en eski kredi kullanım tarihi, kredi kullandıkları tarih boyunca şekillenen en yüksek ve en düşük kredi skorları, kredi notu, ikamet edilen konutun sahipliği, kredinin kullanılması durumunda borçlunun aylık toplam ödeme tutarı, kredi faiz oranı, kredi tutarı, kredinin ödenme durumu (kredinin ödenerek kapatılması veya temerrüde düşmesi), açık kredi sayısı, iflas kayıtları, toplam rotatif kredi bakiyesi ve kredi vadesi dikkate alınarak çalışmamız tamamlanmıştır.

Genel olarak deęerlendirildięinde alıřmamız kapsamında kullanılan LightGBM, Lojistik Regresyon, Yapay Sinir Aęları, Karar Aęaları ve K-En Yakın Komřu algoritmalarının birbirine yakın sonular elde etmiř, te yandan LightGBM algoritmasının %81,04 doęruluk oranı ile ne ıkmıřtır. Kullanılan veri seti zerinden yapılan analiz neticesinde, LightGBM algoritması tarafından;

- temerrde dřmeyeceęi ngrlen 217.689 kredinin temerrde dřmedięi,
- temerrde dřeceęi tahmininde bulunulan 3.338 kredinin temerrde dřtę,
- temerrde dřmeyeceęi ngrlen 48.662 kredinin temerrde dřtę,
- temerrte dřeceęi ngrlen 3.020 kredinin temerrde dřmedięi

belirlenmiřtir.

Anılan veriler ıřıęında, makine ęrenmesi yntemlerinin temerrde dřme olasılıęı bulunan kredilerin tahmininde kayda deęer bařarı saęladıęı grř edinilmiřtir. Dięer taraftan, sz konusu yntemlerin temerrde dřeceęi ngrsnde bulunduęu fakat temerrde dřmeyen krediler nedeniyle pazarlama faaliyetlerinde olumsuzluęa yol aabildięi de tespitlerimiz arasındadır. Bu nedenle, sz konusu algoritmaların hatalı tahminde bulunması durumunda, iřletmeleri stratejik alma ařamalarında yanlıř ynlendirebileceęi hususunun gz ardı edilmemesi gerekmektedir. Bu nedenle, bilhassa bankaların, yalnızca makine ęrenmesi algoritmalarını kullanarak karar almak yerine sz konusu algoritmalarından elde edilecek verileri iř srelerinde destekleyici parametreler olarak dikkate almalarının yerinde olacaęı grřne ulařılmıřtır.

Bir btn olarak deęerlendirdięinde, veri madencilięi ve makine ęrenmesi metotlarının karar alma ařamasında nemli girdiler saęlayabileceęi sonucuna ulařılmıřtır.

KAYNAKÇA

- [1] MALİK Y. (2009), “Enformasyon ve Bilgi Kavramları Bağlamında Enformasyon Yönetimi ve Bilgi Yönetimi”, *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi, Dil ve Tarih-Coğrafya Fakültesi Dergisi*, Cilt 49, Sayı 1, ss. 95-118.
- [2] YILDIZ A. (2014), *İş Zekâsı ve Veri Madenciliğine Dayalı Karar Verme*, Birinci Basım, Gazi Kitabevi, Ankara, ss. 7, 13-14, 39-40, 66-67, 75.
- [3] ŞEKER Ş. E. (2013), *İş Zekâsı ve Veri Madenciliği*, Birinci Basım, Cinius Yayınları, İstanbul, ss. 22, 25, 135, 156.
- [4] PRYTHERCH R. (2005), *Harrod's Librarians' Glossary*, 10th Edition, Ashgate Publishing, England, ss. 195.
- [5] KALSETH K. ve CUMMINGS S. (2001), “Knowledge Management: Development Strategy or Business Strategy?”, *Information Development*, Cilt 17, Sayı 3, ss.163-172.
- [6] KÖSE İ. (2018), *Veri Madenciliği Teori Uygulama ve Felsefesi*, Birinci Basım, Papatya Yayıncılık, İstanbul, ss. 21-22, 48.
- [7] SİLAHTAROĞLU G. (2013), *Veri Madenciliği Kavram ve Algoritmaları*, Üçüncü Basım, Papatya Yayıncılık, İstanbul, ss. 10, 47, 66-67.
- [8] ŞİMŞEK GÜRİSOY U. T. (2012), *Uygulamalı Veri Madenciliği: Sektörel Analizler*, İkinci Basım, Pegem Akademi, Ankara, ss. 2.
- [9] KÖKTÜRK M. S. ve DIRSEHAN T. (2012), *Veri Madenciliği ile Pazarlama Etkileşimi*, Birinci Basım, Nobel Yayıncılık, Ankara, ss. 1.
- [10] GARTNER, *Gartner Glossary Data Mining*, www.gartner.com/en/information-technology/glossary/data-mining, ET. 12.07.2023.
- [11] ŞENTÜRK A. (2006), *Veri Madenciliği Kavram ve Teknikler*, Birinci Basım, Ekin Yayınevi, Bursa, ss.12.
- [12] SAVAŞ S., TOPALOĞLU N. ve YILMAZ M. (2012), “Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri”, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Cilt 11, Sayı 21, ss. 1-23.

- [13]TÜZÜNTÜRK S. (2010), “Veri Madenciliği ve İstatistik”, *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 29, Sayı 1, ss.65-90.
- [14]OĞUZLAR A. (2004), “Veri Madenciliğinde Birliktelik Kuralları”, *Öneri Dergisi*, Cilt 6, Sayı 22, ss.315-321.
- [15]ARSLANTEKİN S. (2003), “Veri Madenciliği ve Bilgi Merkezleri”, *Türk Kütüphaneciliği*, Cilt 17, Sayı 4, ss. 369-380.
- [16]ÇELİK E., DAL D. ve AYDIN T. (2021), “Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması”, *Avrupa Bilim ve Teknoloji Dergisi*, Sayı 27, ss. 880-889.
- [17]HAN J., KAMBER M. ve PEI J. (2012), *Data Mining Concepts and Techniques*, 3th Edition, Morgan Kaufmann Publishers, USA, ss. 3, 23, 26, 328.
- [18]ZHOU Z. (2021), *Machine Learning*, Springer Nature Publisher, Singapore, ss.2.
- [19]MANESH B. (2020), “Machine Learning Algoritihms – A Review”, *International Journal of Science and Research (IJSR)*, Cilt 9, Sayı 1, ss. 381-386.
- [20]SHEARER C. (2000), “The Crisp-DM Model: The New Blueprint for Data Mining”, *Journal of Data Warehousing*, Cilt 5, Sayı 4, ss. 13-23.
- [21]Oracle (2023), *Makine Öğrenimi Nedir?*, www.oracle.com/tr/artificial-intelligence/machine-learning/what-is-machine-learning, ET. 28.07.2023.
- [22]ERSÖZ F. (2013), *Veri Madenciliği ve Uygulamaları*, Birinci Basım, Sage Yayıncılık, Ankara, ss.45.
- [23]ALBAYRAK A. S. ve KOLTAN YILMAZ Ş. (2009), “Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 14, Sayı 1, ss. 31-52.
- [24]EMEL G. G. ve TAŞKIN Ç. (2005), “Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması”, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, Cilt 6, Sayı 2, ss. 221-239.
- [25]YALÇIN Ö. (2008), *Veri Madenciliği Yöntemleri*, Birinci Basım, Papatya Yayıncılık, İstanbul, ss.118.
- [26]BHATIA N. ve VANDANA (2010), “Survey of Nearest Neighbor Techniques”, *International Journal of Computer Science and Information Security*, Cilt 8, Sayı 2, ss. 302-305.
- [27]BLUM A. (1992), *Neural Networks in C++ An Object-Oriented Framework For Building Connectionist Systems*, John Wiley&Sons, USA.

- [28]ŞEN Z. (2004), *Yapay Sinir Ağları*, Birinci Basım, Su Kitapevi, İstanbul, ss.4.
- [29]ÖZTEMEL E (2006), *Yapay Sinir Ağları*, İkinci Baskı, Papatya Yayıncılık, İstanbul.
- [30]ÖZEKES S. (2003), “Veri Madenciliği Modelleri ve Uygulama Alanları”, *İstanbul Ticaret Üniversitesi Dergisi*, Cilt 2, Sayı 3, ss.65-82.
- [31]BDDK (2023), Bankacılık Düzenleme ve Denetleme Kurulu, *Türk Bankacılık Sektörü Temel Göstergeleri*, www.bddk.org.tr/Veri/EkGetir/8?ekId=160, ET. 31.07.2023.
- [32]BENLİĞİRAY Y. ve BANAR K. (2006), *Banka ve Sigorta Hukuku*, Anadolu Üniversitesi, Birinci Basım, Anadolu Üniversitesi Yayınları, Eskişehir.
- [33]AKGÜÇ Ö. (2006), *Kredi Taleplerinin Değerlendirilmesi*, Yedinci Basım, Arayış Basım ve Yayıncılık, İstanbul.
- [34] BDDK (2023), Bankacılık Düzenleme ve Denetleme Kurulu, *Aylık Bankacılık Sektörü Verileri*, www.bddk.org.tr/Veri/Detay/159, ET. 27.07.2023.
- [35]KE G., MENG Q., FINLEY T., WANG T., CHEN W., MA W. ve LIU T. Y. (2017), "Light GBM: A Highly Efficient Gradient Boosting Decision Tree", *Advances in Neural Information Processing Systems (NIPS)*, Cilt 30, ss. 1-8.
- [36]TANINMIŞ YÜCEMEMİŞ B. ve SÖZER İ. (2011), “Bankalarda Takipteki Krediler: Türk Bankacılık Sektöründe Takipteki Kredilerin Tahminine Yönelik Bir Model Uygulaması”, *Finansal Araştırmalar ve Çalışmalar Dergisi*, Cilt 3, Sayı 5, ss. 43-56.
- [37]SELİMLER H. (2015), “Sorunlu Kredilerin Analizi, Banka Finansal Tablo ve Oranlarına Etkisinin Değerlendirilmesi”, *Finansal Araştırmalar ve Çalışmalar Dergisi*, Cilt 7, Sayı 12, ss.131-172.
- [38]MAYANK A., ARUN V. ve PAWAN W. (2022), “Prediction of Loan Behaviour with Machine Learning Models for Secure Banking”, *Journal of Computer Science an Engineering (JCSE)*, Cilt 3, Sayı 1, ss.1-13.
- [39]BDDK (2014), Bankacılık Düzenleme ve Denetleme Kurulu, *Bankaların İç Sistemleri ve İçsel Sermaye Yeterliliği Değerlendirme Süreci Hakkında Yönetmelik*, Ankara.
- [40]TBB (2023), Türkiye Bankalar Birliği, *Risk Yönetimi*, https://www.tbb.org.tr/Dosyalar/Arastirma_ve_Raporlar/riskyonetimi.pdf, ET. 24.07.2023.

- [41]KKB (2023), Kredi Kayıt Bürosu, *Hakkımızda*, www.kkb.com.tr/hakkimizda, ET. 17.07.2023.
- [42]FİNDEKS, *Ürünler*, www.findeks.com/urunler/findeks-kredi-notu, ET. 11.07.2023.
- [43]BDDK (2023), Bankacılık Düzenleme ve Denetleme Kurulu, *İyi Uygulama Rehberi*, <https://www.bddk.org.tr/Mevzuat/DokumanGetir/1041>, ET.10.07.2023.
- [44]Financial Stability Board (2017), *Artificial Intelligence and Machine Learning In Financial Services*, İsviçre.
- [45]CARR B., BAILEY N., FERENZY D. (2019), “Machine Learning In Credit Risk Report”, *Institute of International Finance*, A.B.D.
- [46]MARTIN L., SHARMA S. ve MADDULETY K. (2019), “Machine Learning in Banking Risk Management: A Literature Review”, *Risks*, Cilt 7, Sayı 1, ss. 29.
- [47]BALAN V. J. R. V. S. (2013), “A Review On Data Mining In Banking Sector”, *American Journal of Applied Sciences*, Cilt 10, Sayı 10, ss. 1160-1165.
- [48]ÇİĞŞAR B. ve ÜNAL D. (2019), “Comparison of Data Mining Classification Algorithms Determining the Default Risk”, *Scientific Programming*, Cilt 2019, ss.1-2.