

Kredi Tahsisinde Kredi Temerrüdü Tahmini Modeli

Araştırma Makalesi/Research Article

 Çiğdem TİMAS¹,  Nuri BİNGÖL²

¹ Fen Bilimleri Enstitüsü, Yapay Zeka Mühendisliği, Üsküdar Üniversitesi, İstanbul, Türkiye

² Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, Üsküdar Üniversitesi, İstanbul, Türkiye

cigdem.timas1981@gmail.com, nuri.bingol@uskudar.edu.tr

(Geliş/Received:11.11.2024; Kabul/Accepted:07.03.2025)

DOI: 10.17671/gazibtd.1583269

Özet— Finans sektörünün ana öncüleri olan bankalar ve diğer finans kurumları piyasayı kullandırmış oldukları krediler sayesinde fonlamaktadırlar. Böylelikle fonlamanın sağlıklı yürütülebilmesi, fonlamanın en büyük kalemi olan kredi kullandırma işlevinin doğru yapılması ile sağlanır. Bu da ancak ve ancak tahsis edilen kredilerin bir kısmının ya da tamamının zamanında geri dönmeme olasılığını belirten “kredi riski” durumunun iyi idare edilebilmesi yani geri dönmeyen kredilerin mümkün mertebe en düşük seviyeye indirilmesi ile sağlanabilmektedir. Veri kümesine **Karar Ağacı**, **CatBoost**, **Yapay Sinir Ağları (YSA)** ve **Random Forest** ve karşılaştırma imkânı elde etme açısından da geleneksel istatistiki yöntemlerden **Lojistik Regresyon (LR)** yöntemleri uygulanarak kredi temerrüdü tahminleri yapılmıştır. (Özet)

Anahtar Kelimeler— temerrüd, tahmin modeli ,kredi tahsisi

Credit Default Prediction Model in Credit Allocation

Abstract— The main pioneers of the finance sector, banks and other financial institutions, fund the market through the loans they provide. Thus, the healthy execution of funding is ensured by correctly performing the loan provision function, which is the largest item of funding. This can only be achieved by managing the “credit risk” situation, which indicates the possibility of not returning some or all of the allocated loans on time, in other words, by reducing the non-performing loans to the lowest possible level. Credit default estimates were made by applying **Decision Tree**, **CatBoost**, **Artificial Neural Networks (ANN)** and **Random Forest** to the data set and **Logistic Regression (LR)** methods, which are among the traditional statistical methods, for comparison purposes.(Abstract)

Keywords— default, prediction model, credit allocation

1. GİRİŞ (INTRODUCTION)

Bankalar tarafından tahsis edilen kredilerin, müşteri tarafından geri ödenmemesi veya zamanında geri ödenmemesi hem kredi veren işletmenin sermaye kaybını hem de genel ekonomide oluşabilecek çeşitli riskli durumları beraberinde getirmektedir. En başta Bankaların yer aldığı finans dünyası, ekonomilerin devamlılığı açısından stratejik ve çok önemli bir işlev olan, “fonlama” işlevini yerine getiren, kritik derece de öneme sahip bir sektör niteliğindedir. Finans sektörünün ana öncülleri olan bankalar, diğer finans kurumları piyasayı kullandırmış oldukları krediler sayesinde fonlamaktadırlar. Böylelikle fonlamanın sağlıklı yürütülebilmesi, fonlamanın en büyük kalemi olan kredi kullandırma işlevinin doğru yapılması ile sağlanır. Bu da ancak ve ancak tahsis edilen kredilerin bir kısmının ya da tamamının zamanında geri dönmeme olasılığını belirten “kredi riski” durumunun iyi idare edilebilmesi yani geri dönmeyen kredilerin mümkün mertebe en düşük seviyeye indirilmesi sağlanabilmektedir. Kredi başvurusunda bulunan bu müşterilerden hangilerinin riskli kategorisine girmeyip kredi başvurusunun olumlu yanıt dönülmesi ve kredinin tahsis edilmesi, hangilerinin ise riskli bulunup taleplerinin reddedilmesi gerektiği anlaşılmaya çalışılmıştır. Belirtildiği gibi bu sorun, müşterileri “riskli” ve “risksiz” şeklinde iki sınıfa ayırma, yani bir çeşit sınıflandırma problemidir. Çalışmada veri kümesine Karar Ağacı, CatBoost, Yapay Sinir Ağları ve Random Forest ve karşılaştırma imkânı elde etme açısından da geleneksel istatistiki yöntemlerden Lojistik Regresyon yöntemleri uygulanmıştır. Birden fazla makine öğrenmesi modeli test edilmiş ve performansları karşılaştırılmıştır. Kredi temerrüdü tahmin modelinin performansını değerlendirmek için Confusion Matrix, ROC-AUC skoru ve Classification Report metriklerini kullandık. Confusion Matrix, modelin doğru ve yanlış tahminlerini analiz ederek False Negative (FN) ve False Positive (FP) değerlerini inceledik. ROC-AUC skoru, modelin sınıflandırma başarısını ölçerek temerrüde düşenleri ne kadar iyi ayırt ettiğini değerlendirdik. Ayrıca, Precision (Kesinlik), Recall (Duyarlılık) ve F1-Score gibi sınıflandırma metrikleriyle modelin yanlış pozitif ve yanlış negatif tahminleri arasındaki dengesini analiz ettik. Accuracy tek başına yeterli olmadığından, özellikle Recall ve AUC-ROC değerlerini karşılaştırarak en iyi modelin bulunması hedeflenmiştir. Çalışma makine öğrenmesiyle kredi risk tahmininde geleneksel lojistik regresyon yerine CatBoost, Random Forest gibi güçlü modeller de kullanılarak, literatürdeki klasik yöntemlere kıyasla bu modellerin de başarılı performanslar sunabileceğini göstermesi yönünden önemlidir.

2. LİTERATÜR (LITERATURE REVIEW)

Literatürde kredi riski tahminine yönelik çeşitli çalışmalar yapılmıştır. Bunlardan biri olan Yeh [1], Tayvan'daki kredi kartı kullanıcılarının temerrüt etme olasılığını, altı farklı veri madenciliği yöntemi ile analiz ederek bu yöntemlerin doğruluk oranlarını karşılaştırmıştır. Bu çalışmada değerlendirilen yöntemler; diskriminant analizi, lojistik

regresyon, Bayes sınıflandırıcısı, en yakın komşu algoritması, yapay sinir ağları ve sınıflandırma ağaçlarıdır. Elde edilen sonuçlara göre, bu teknikler arasında en iyi performansı yapay sinir ağları göstermiş ve kredi puanlama uygulamalarında etkili bir yöntem olarak öne çıkmıştır. [1]

Budak ve Erpolat çalışmalarında [2], bankaların kredi risklerini en aza indirmek amacıyla, kredi başvurusunda bulunan müşterilerin ödeme potansiyellerini analiz etmek için yapay sinir ağları ve lojistik regresyon analizi karşılaştırılmıştır. Çalışmanın sonuçlarına göre, yapay sinir ağları yöntemi, müşterilerin ödeme alışkanlıklarının düzenli olup olmadığını tahmin etmede lojistik regresyon modeline kıyasla daha yüksek bir başarı sergilemiştir. [2]

Kavcıoğlu [3]'ün çalışmasında, kredi riskinin ölçümünde kullanılan modellerin genel olarak birbirine benzediği, ancak farklı değerlendirme kriterleri ve ölçüm yöntemleri içerdiği vurgulanmıştır. Bu modellerin bazı yönlerden birbirlerine avantaj sağladığı, ancak bazı eksikliklerinin de bulunduğu ifade edilmiştir. Bu sebeplerle, özellikle bankalar gibi kredi veren kurumların risklerini en aza indirebilmeleri için kendi yapılarına en uygun modeli ve yöntemi seçmeleri gerektiği belirtilmiştir. [3]

Hamori ve çalışma arkadaşları [4], Tayvan'daki temerrüt ödeme verilerini analiz ederek üç toplu öğrenme yöntemi (torbalama, rastgele orman ve artırma) ile çeşitli sinir ağı yöntemlerinin tahmin doğruluğu ve sınıflandırma yeteneğini karşılaştırmıştır. Çalışmanın sonuçlarına göre, güçlendirme yönteminin sınıflandırma başarısının, sinir ağları da dahil olmak üzere diğer makine öğrenimi yöntemlerine göre daha yüksek olduğu ortaya çıkmıştır. Ayrıca sinir ağı modellerinin performansının, seçilen aktivasyon fonksiyonuna ve gizli katman sayısına göre değişiklik gösterdiği bulunmuştur. [4]

İldeş [5]'in çalışmasında, kredi riskinin ölçümüne yönelik birçok farklı yöntemin mevcut olduğu, klasik ve modern kredi risk ölçüm metodlarının yanı sıra, portföy kredi riskinin değerlendirilmesi amacıyla uluslararası finansal kuruluşlar tarafından geliştirilen modellerin de kullanılmasının önemli olduğu vurgulanmıştır. [5]

Mahbobi, Kimiagari ve Vasudevan [6] çalışmalarında, entegre bir tahmin doğruluğu algoritması geliştirirken, çalışmada makine öğrenimi sınıflandırıcıları olarak DNN (Derin Sinir Ağı), SVM (Destek Vektör Makineleri), KNN (En Yakın Komşu) ve ANN (Yapay Sinir Ağı) kullanılmıştır. 30.000 dengesiz veri kümesinden yararlanılarak, temerrüt ödemelerinin tahmin doğruluğunu artırmak amacıyla Sentetik Azınlık Aşırı Örneklemme Tekniği (SMOTE), SVM SMOTE, rastgele düşük örneklemme ve ALL-KNN gibi çeşitli aşırı ve yetersiz örneklemme stratejileri uygulanmıştır. Çalışmanın sonuçlarına göre, ALL-KNN örneklemme tekniği kullanılarak yapılan analizde, SVM modelinin %98,6 doğruluk ve en düşük çapraz entropi kaybı değeri olan 0,028 ile en iyi performansı sergilediği gözlemlenmiştir. [6]

Tütüncü ve Gürsakar [7], çalışmalarında kredi skorlama sistemlerinde en başarılı tahmini yapan algoritmayı belirlemek amacıyla makine öğrenmesi yöntemlerini kullanmışlardır. Gradyan Artırma, Yapay Sinir Ağları, Lojistik Regresyon, Rastgele Orman, Karar Ağacı, Destek Vektör Makineleri, K-En Yakın Komşu ve WOE dönüşümleriyle Lojistik Regresyon algoritmaları için modeller oluşturmuşlardır. Çalışmanın sonuçlarına göre, temerrüde düşen ve düşmeyen müşterilerin en iyi sınıflandırmasını sağlayan algoritmanın Gradyan Artırma olduğu tespit edilmiştir. [7]

Milli [8], tez çalışmasında riskli müşterilerin belirlenmesi için K-En Yakın Komşu, Naive Bayes, Lojistik Regresyon, Destek Vektör Makineleri, Çok Katmanlı Algılayıcı, Karar Ağaçları, Rastgele Ormanlar, Gradyan Artırma Karar Ağaçları, Extra Ağaçlar, Sert ve Yumuşak Oylama gibi çeşitli makine öğrenmesi sınıflandırma tekniklerini kullanmıştır. Ayrıca sınıf dengesizliği sorununu çözmek amacıyla, RUS (Random UnderSampling), ROS (Random OverSampling), SMOTE-ENN, SMOTE-Tomek Links ve Balanced Bagging Classifier gibi yeniden örnekleme ve hibrit yöntemlerle veri dengesini sağlamıştır. Üç farklı kredi veri kümesi üzerinde dört farklı senaryo ile yapılan çalışma sonucunda, sınıf dengesizliği için alt ve üst örnekleme yöntemlerinin bir arada kullanıldığı hibrit yöntemin, makine öğrenmesi tekniklerinde en iyi sınıflandırma performansına ulaşmada etkili olabileceği sonucuna varılmıştır. [8]

Zhu ve çalışma arkadaşlarının [9] yaptığı çalışmada, kredi temerrüdünü tahmin etmek amacıyla lojistik regresyon, karar ağacı, XGBoost ve LightGBM modelleri kullanılmıştır. Elde edilen sonuçlar, LightGBM ve XGBoost'un tahmin performansının lojistik regresyon ve karar ağacı modellerine göre daha yüksek olduğunu göstermiştir. LightGBM modeli için eğri altındaki alan 0,7213 olarak bulunmuş ve her iki modelin doğruluğu 0,8'in üzerinde çıkmıştır. Ayrıca, LightGBM ve XGBoost'un hassasiyetinin 0,55'in üzerinde olduğu belirlenmiştir. Sonuçlara göre, kredi vadesi, kredi notu, kredi ratingi ve kredi tutarı gibi değişkenlerin tahmin sonuçları üzerinde etkili olduğu görülmüştür. [9]

Literatür incelendiğinde, farklı makine öğrenmesi algoritmalarının kredi riskinin ölçümünde en iyi sonuçları verdiğine dair genel bir fikir birliği olmadığı görülmektedir. Bu çalışmayla Karar Ağacı ,CatBoost ,Yapay Sinir Ağları,Random Forest ve Lojistik Regresyon yöntemleriyle sınıflandırma performansları karşılaştırılarak kredi riski ölçümünde en iyi sonucu veren yöntemin bulunması hedeflenmiştir.

3. YÖNTEM (METHOD)

Veri kümesi Lending Club'a ait açık erişimde olan 27 değişkenden 396030 gözlemden oluşan veri setidir. Veri kümesi geçmişte kredi başvurusunda bulunanlar ve bunların temerrüt edip etmediğine ilişkin bilgileri içermektedir. Bu tarz modellerin kullanılmasındaki amaç, tekrardan kredi kullanıldığında bir kişinin temerrüde

düşme ihtimalinin olup olmadığını tespit edip ya krediyi reddetmek, ya da talep edilen kredi miktarından daha az tutarda bir kredi onaylamak, ya da riskli müşterilere daha yüksek faiz oranıyla borç vermek ya da direkt başvuruyu reddetmek şeklinde olabilir.

Araştırma kapsamında, Python'un sunduğu scikit-learn, pandas, numpy, matplotlib ve seaborn gibi güçlü kütüphanelerden faydalanılarak, geniş çaplı veri analizi ve makine öğrenmesi algoritmaları uygulanmıştır. Ham verinin temizlenmesi ve eksik değerlerin uygun yöntemlerle doldurulması sağlanmış, ardından değişkenlerin istatistiksel dağılımları detaylı bir şekilde incelenmiştir. Korelasyon analizleri ve değişken seçimi teknikleri kullanılarak modele en uygun özellikler belirlenmiş, böylece tahmin gücü yüksek ve doğruluğu artırılmış bir model oluşturulmuştur.

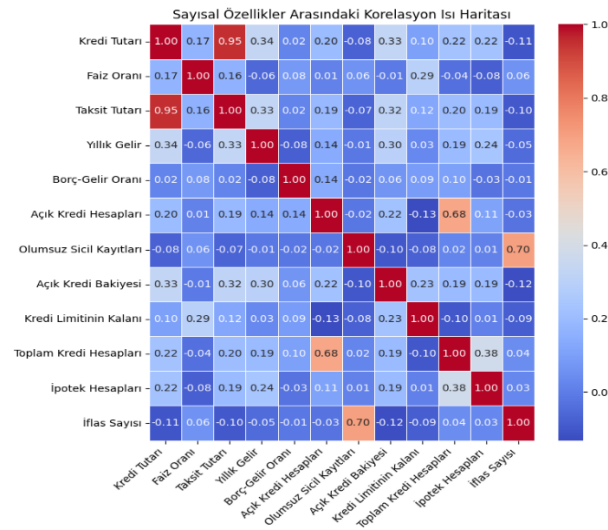
Araştırmada, lojistik regresyon, karar ağaçları, rastgele ormanlar ve yapay sinir ağları gibi farklı makine öğrenmesi algoritmaları karşılaştırılmış ve performans ölçütleri dikkate alınarak en iyi sonuç veren model seçilmiştir. Modelin başarı oranını değerlendirmek amacıyla, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi çeşitli metrikler hesaplanmış, ayrıca ROC eğrisi ve AUC skoru ile modelin genel performansı görselleştirilmiştir.

Python'un esnek ve güçlü analiz araçları kullanılarak yapılan bu modelleme süreci, yalnızca veri setinden anlamlı içgörüler çıkarmakla kalmamış, aynı zamanda geleceğe yönelik doğru tahminlerin yapılmasını mümkün kılmıştır. Araştırmada geliştirilen bu metodoloji, veri bilimi alanında çalışan araştırmacılar ve sektör profesyonelleri için yol gösterici bir rehber niteliği taşımaktadır.

4. BULGULAR (FINDINGS)

4.1 Veri Analizi (Data Analysis)

Veri kümesinde kredi modelinin belirlenebilmesi için birçok değişken bulunmaktadır. Kredi faiz oranı, kredi miktarı, kredi notu, borçlunun kamu kayıtlarındaki iflas geçmişi, meslek bilgisi ve ev sahipliği durumu gibi çeşitli değişkenler yer almaktadır. Sayısal değerler arasındaki korelasyonu analiz etmek amacıyla ısı haritası kullanılmış ve kredi tutarı ile taksit değerleri arasında 0.95 gibi yüksek bir pozitif korelasyon katsayısı olduğu tespit edilmiştir.



Şekil 1. Sayısal Özellikler Arasındaki Korelasyon Isı Haritası (Correlation Heat Map Between Numerical Features)

Veri setinde eksik değerlerin (missing values) dağılımına bakıldığında meslek bilgisi, meslekte geçen süre, mortgage hesapları değişkenlerinde fazlaca sayıda eksik bilgi olduğu görülmüştür. Çok sayıda iş unvanının bulunması, iş unvanı değişkenindeki eksik verilerin sayıca fazla olması (22927) ve değişkenin modele katkısının çok olmayacağını değerlendirilerek veri setinden çıkarılmıştır. Aynı pozisyondaki geçen yıl sayısını gösteren bir fonksiyon döngüsü yazdırılarak yüzdelik oranları incelenmiş hemen hemen hepsinin yakın değerler olduğu analiz edilmiştir. Meslekte geçen çalışma sürelerinin de modele katkı sağlamayacağı değerlendirilerek veri setinden çıkarılmıştır. Kredi başvuru amacı değişkeni incelendiğinde kredi temerrüdü tahmini modelinin amacına uygun bir girdi sağlamadığı değerlendirilmiştir. Mortgage hesapları kredi için başvuran adayın ipotek hesaplarının sayısını gösteren bir değişken olup bu değişkendeki eksik değerler (37795) de sayıca fazladır. Mortgage hesapları değişkeni, kamu kayıtlarındaki iflas değişkeni ve kamu kayıtları değişkenindeki eksik değerler için fonksiyon oluşturulmuş ve fonksiyonla ilgili değişkenlerin kategorik (binary - ikili formata) hale getirilmesi sağlanmıştır. Kamu Kayıtlarındaki İflas değişkeni Sabit Değer ile Doldurma (Constant Imputation) yöntemi ile eksik değerlere sıfır atanarak doldurulmuştur. Ortalama ile Doldurma (Mean Imputation) yöntemiyle de Mortgage hesapları değişkenindeki eksik veriler doldurulmuştur. Kredi başvuru adayının adres değişkeni, kredinin finanse edildiği ay bilgisi ve kredi kulübünün verdiği not değişkenleri veri setinden çıkarılmıştır. Kredi durumu kategorik değişkenleri 0 ve 1 sayısal değerlerine dönüştürülerek modelin işleyebileceği duruma dönüştürüldü. Bağımlı değişken; kredi statüsü (loan_status) değişkeni 1 kredi ödemelerinde gecikmesi (temerrüd) olan, 0 ise kredi ödemelerinde gecikmesi olmayan kredisini zamanında ödeyen ifade etmektedir.

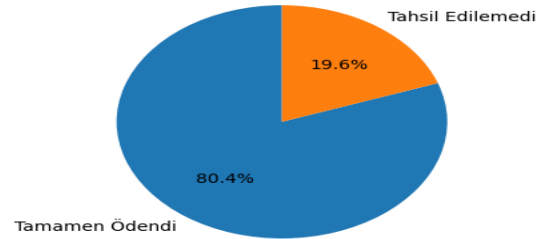
Şekil 1'den Şekil 9'a kadar olan görseller, yazarlar tarafından model oluşturulmadan önce veriyi tanımak ve

modele hazırlık yapmak amacıyla gerçekleştirilen veri analizi çalışmalarına dayanmaktadır.

4.1.1 Kredi Statüsüne Göre Taksit Dağılımı (Distribution of Installment by Credit Status)

Veri kümesinde kredi statülerine göre, 77673 adedinde 19.6%'sında kredisi temerrüde düşmüş (Tahsil edilemedi), 318357 adedinde 80.4 % 'sında kredisi tahsil edilmiş (Tamamen Ödendi) olarak veriler dağılım göstermektedir.

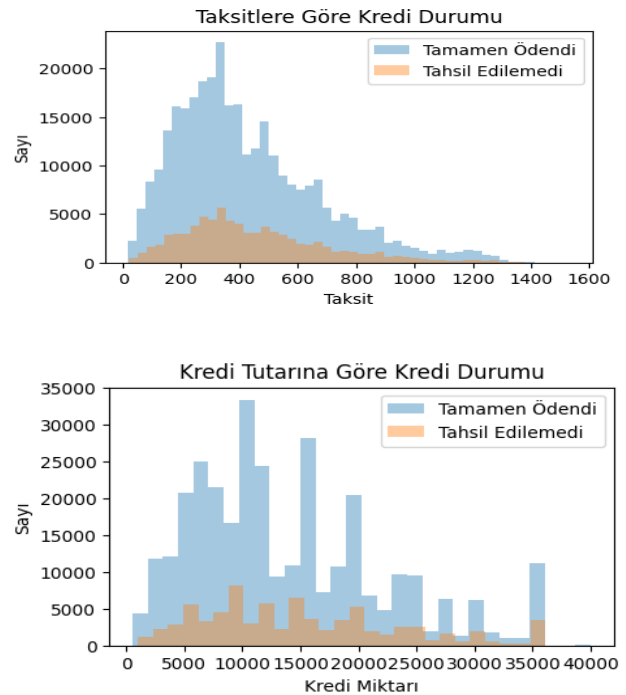
Kredi Statüsüne Göre Taksit Dağılımı



Şekil 2. Kredi Statüsüne Göre Taksit Dağılımı (Installment Distribution According to Loan Status)

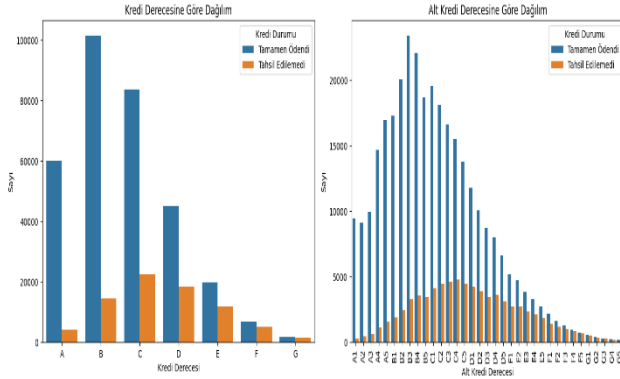
4.1.2 Taksit ve Kredi Tutarı & Kredi Durumu Grafiği (Installment and Loan Amount & Loan Status Chart)

Taksit tutarı ve kredi tutarı değişkenlerine göre kredi durumlarının görüntülenmesi sağlanmıştır. Histogramlarda taksit tutarı ve kredi tutarı arttıkça geri ödeme durumunun azaldığı görülmektedir. Kısa vadede ve daha düşük tutardaki kredilerin daha çok tahsil edilebildiği anlaşılmaktadır.



Şekil 3. Taksit ve Kredi Tutarı & Kredi Durumu Grafiği (Installment and Loan Amount and Loan Status Graph)

4.1.3 Derece(grade) ve Alt derece(sub_grade) Kredi Geri Ödeme Durumuna Göre Dağılımı (Distribution of Grade and Sub-grade by Loan Repayment Status)



Şekil 4. Derece ve Alt derecenin Kredi Geri Ödeme Durumuna Göre Dağılımı (Distribution of Degree and Sub-degree According to Loan Repayment Status)

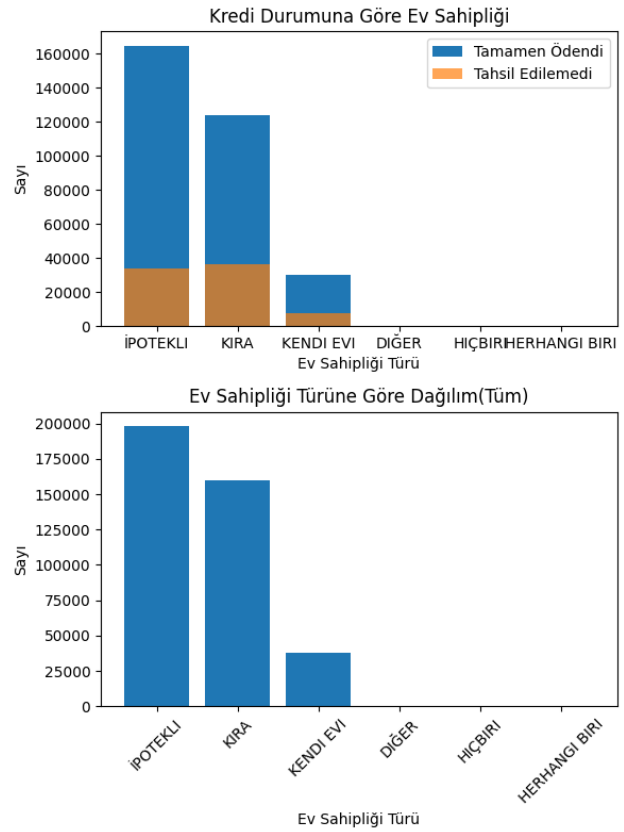
Derece: Borçlunun genel kredi güvenilirliğini veya risk seviyesini gösteren geniş bir kategoridir. Genellikle bir finans kurumu veya kredi derecelendirme kuruluşu tarafından borçlunun kredi skoru, finansal geçmişi ve diğer faktörlere dayanarak atanır. Harflerle ifade edilir (örneğin, A, B, C), burada "A", yüksek kredi skoruna sahip, düşük riskli bir borçluyu temsil ederken, "C" ve altı, daha yüksek riskli borçluları ifade eder. Yani, borçlunun krediyi geri ödeme olasılığını belirlemek için kullanılır.

Alt Derece: Derece içinde daha ayrıntılı bir sınıflandırmadır. Aynı derecedeki borçlular arasındaki risk farklarını daha net bir şekilde ayırt eder. Örneğin, "A" derecesi altında "A1", "A2", "A3" gibi alt dereceler olabilir. "A1", o kategorideki en güvenilir borçluyu temsil ederken, "A3" biraz daha fazla risk taşıyan borçluyu ifade eder. Bu alt dereceler, aynı kredi derecesine sahip borçlular arasında bile risk farklılıklarını ortaya koyar.

Derece ve altderece sütunlarındaki verilerin kredi ödeme durumu değişkenine göre nasıl dağıldığını gösterir. F ve G alt sınıflarının düzenli olarak geri ödenmediği görülmekle birlikte, A, B ve C sınıflarının geri ödeme durumlarının başarılı olduğu görülmektedir.

4.1.4. Ev Sahipliği Durumu ile Kredi Durumu Arasındaki İlişki (The Relationship Between Home Ownership Status and Loan Status)

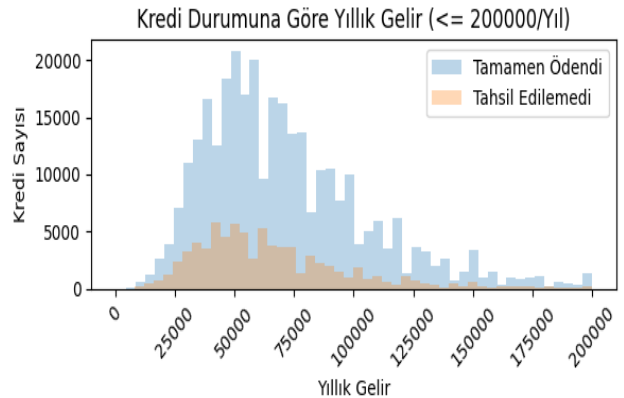
Kredi durumu ve ev sahipliği durumu değişkenleri arasındaki ilişkiyi görsel olarak incelemek ve ev sahipliği durumlarına göre kredi durumlarını karşılaştırmak amaçlanmıştır. Tamamı ödenmiş ve Ödenmemiş borç (Tahsil edilemedi) durumlarına sahip olanlar arasında ev sahipliği durumlarının dağılımına bakıldığında, Tamamının ödeme durumunun ipotekli (Mortgage) ve kirada ikamet edenler arasında yoğunlaştığı görülmektedir.



Şekil 5. Ev Sahipliği Durumu ile Kredi Durumu Arasındaki İlişki (The Relationship Between Homeownership Status and Loan Status)

4.1.5. Yıllık Geliri <200.000 \$ Olanlarda Kredi Ödeme Durumu (Loan Payment Status for Annual Income < \$200,000)

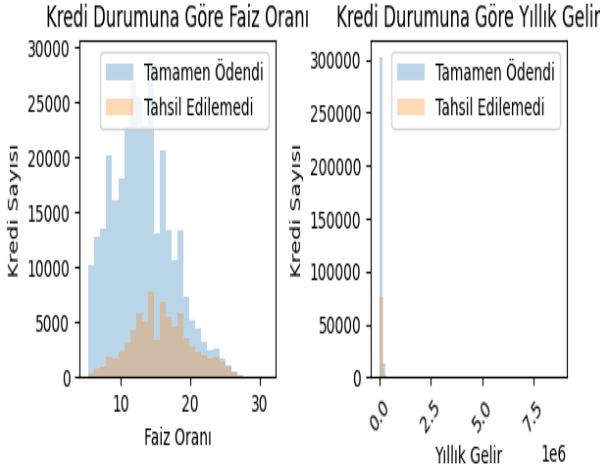
Yıllık geliri 200.000 \$ veya daha az olan verilere göre kredi ödeme durumu kategorisine göre gruplanıp bir histogram grafiği oluşturuldu. Grafiğe bakıldığında büyük dağılımın yıllık geliri 100.000 \$ ve altında olanlarda olduğu görülmektedir.



Şekil 6. Yıllık Geliri <200.000 \$ Olanlarda Kredi Ödeme Durumu (Loan Payment Status for Annual Income < \$200,000)

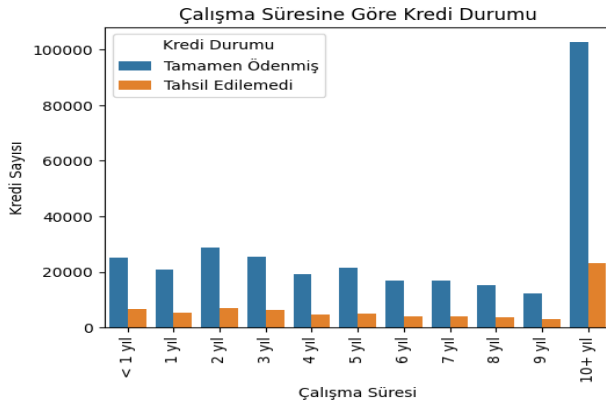
4.1.6 Faiz Oranı ile Yıllık Gelir ve Kredi Durumu Arasındaki İlişki (The Relationship Between Interest Rate, Annual Income, and Loan Status)

Faiz oranı ve yıllık gelir nitelikleri kredi durum değişkenine göre gruplanarak bir histogram çizdirilmesi sağlandı. Histograma göre faiz oranı arttıkça, kredi kullandırımının azaldığı ve temerrüd oranının arttığı tespit edilmiştir.



Şekil 7. Faiz Oranı ile Yıllık Gelir ve Kredi Durumu Arasındaki İlişki (The Relationship Between Interest Rate, Annual Income, and Loan Status)

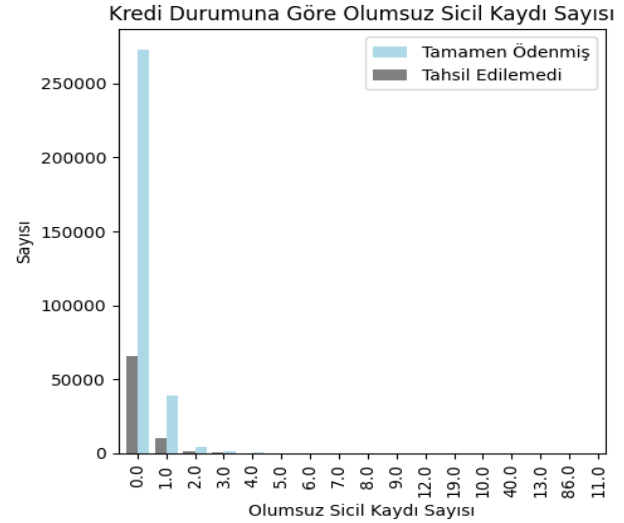
4.1.7 Meslekteki Tecrübe ve Kredi Durumu İlişkisi (Relationship Between Professional Experience and Loan Status)



Şekil 8. Meslekteki Tecrübe ve Kredi Durumu İlişkisi (Relationship Between Professional Experience and Loan Status)

Kredi başvurularında 10 yıl ve üzeri mesleki deneyime sahip bireylerin geri ödeme performansında belirgin bir pozitif farklılık gözlemlenmiştir.

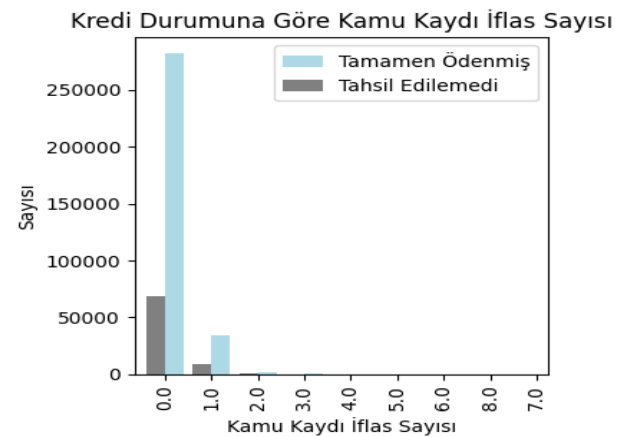
4.1.8 Olumsuz Sicil Kaydı ve Kredi Durumları (Negative Credit Record and Loan Status)



Şekil 9. Olumsuz Sicil Kaydı ve Kredi Durumları (Negative Credit Record and Loan Status)

Olumsuz sicil kaydı, bir borçlunun itibarını olumsuz etkileyen kamuya açık kayıtların sayısını ifade eder. Bu kayıtlar genellikle kredi geçmişi, borçların ödenmesi, icra takibi gibi finansal durumlarıyla ilgili olabilir. Bankaların kredi kullandırım kararı esnasında müşterinin Memzuç kayıtlarına, KKB kayıtlarına bakarak karar vermesi örnek verilebilir. Bu olumsuz kayıtlar genellikle kredi notunun düşük olmasına neden olabilir ve finansal geçmişin incelenmesinde ve kredi tahsis kararının verilmesinde önemli bir etken olabilir. Geçmişte olumsuz sicil kaydı bulunmayanlarda gecikmeye düşme durumu gözle görülür derecede azdır. Fakat yine grafiğe bakıldığında geçmişte olumsuz sicil kaydı olmamasına rağmen, kredisi temerrüde düşenlerin sayısı, daha önce olumsuz kaydı olanlara göre fazladır.

4.1.9 Kamu Kayıtlarındaki İflas Kaydı & Kredi Durumları (Bankruptcy Records in Public Records & Loan Status)

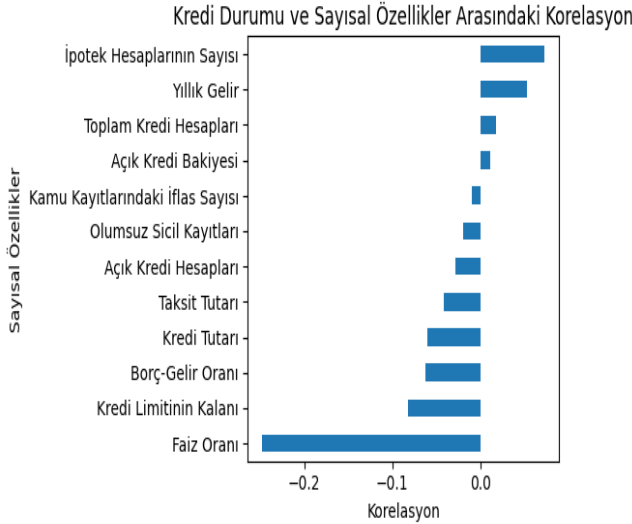


Şekil 10. Kamu Kayıtlarındaki İflas Kaydı & Kredi Durumları (Bankruptcy Records in Public Records & Loan Status)

Genel olarak başvuran kişinin kamu kayıtlarındaki iflas kaydı sayısının yüksek olması kredi geçmişinin olumsuz durumu olarak değerlendirilir ve kredi notunun düşmesine sebep olabilir. Çünkü bu tür iflaslar, kredi verenler için

riskli bir durum olarak kabul edilir ve gelecekteki kredi başvuruları için olumsuz bir etkiye sahip olur. Bu sebeple önemli bir değişkendir. Grafiğe bakıldığında geçmişte kamuda iflas kaydı olmayan müşterilerin kredi ödemelerinde büyük çoğunluğunda sorun olmadığı görülmüştür. Fakat yine grafiğe bakıldığında geçmişte kamuda iflas kaydı olmamasına rağmen ,kredisi temerrüde düşenlerin sayısı ,daha önce olumsuz kaydı olanlara göre fazladır.

4.1.10 Korelasyon - Kredi Durumu ve Sayısal Özellikler(Correlation Between Loan Status and Numerical Characteristics)



Şekil 11.Kredi Durumu ve Sayısal Özellikler Arasındaki Korelasyon(Correlation Between Credit Status and Numerical Characteristics)

Korelasyona bakıldığında faiz oranı ile kredinin ödenme durumu arasında negatif ve güçlü bir korelasyon olduğu görülmektedir. Faiz oranı arttıkça kredinin ödenme durumu da zorlaşmaktadır. İpotek hesaplarının sayısı ve yıllık gelir ile kredi durumları arasında da pozitif bir korelasyon olduğu görülmektedir. Bu değerler 1 e doğru yaklaştığı için aralarında güçlü bir ilişki olduğu görülmektedir.

4.2 Model Geliştirme(Model Development)

Veri analizi sonucunda belirlenen nihai öznitelik değişkenleri kullanılarak kredilerin temerrüt riski aşağıdaki algoritmalar ile belirlenmeye çalışılmış ve kullanılan algoritmaların sınıflandırma performansları karşılaştırılmıştır.

Makine öğrenmesi modeli oluşturmak için veri bölme, ölçeklendirme ve değişken ayırıştırma işlemleri gerçekleştirilmiştir. Veri kümesi %67 eğitim (train) ve %33 test (test) olmak üzere ikiye ayrıldı ve random_state=42 ile sonuçların her seferinde aynı olması için rastgelelik sabitlendi. Bağımlı değişken kredi durumu (loan_status) değişkeni, kredinin ödenip ödenmediğini gösteren hedef değişken olup kredi miktarı, faiz oranı, yıllık gelir,uygunsuz kamu sicil kayıtları ,açık kredi kart bakiyesi

gibi değişkenler bağımsız değişkenler olarak ayrılmıştır. Min-Max Ölçeklendirme (Feature Scaling) gerçekleştirilmiş olup MinMaxScaler() ile tüm bağımsız değişkenleri 0 ile 1 arasına ölçeklendirildi.

Sınıflandırma modelinin tahmin performansını değerlendirmek için Confusion(karışıklık) matrisi kullanıldı.Böylelikle modelin doğru ve yanlış sınıflandırma sayıları gösterildi. Matris, tahminleri daha açık bir şekilde gösterir ve modelin performansının anlaşılmasına yardımcı olur.

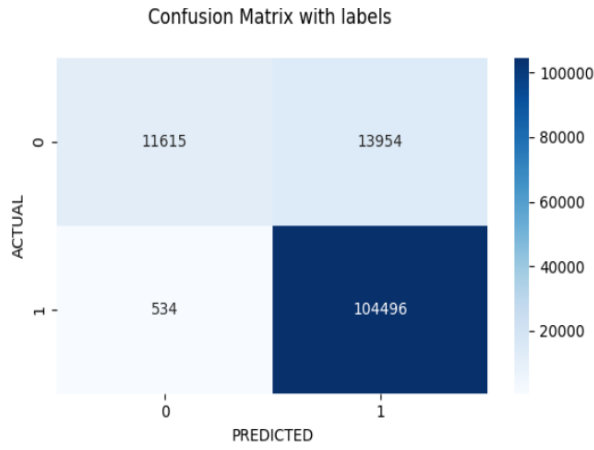
Karışıklık matrisi dört ana hücreye bölünür ve şu dört kategoriye içerir:

1. **True Positive (TP):** Modelin doğru bir şekilde pozitif olarak sınıflandırdığı örnekler.
2. **False Positive (FP):** Modelin yanlış bir şekilde pozitif olarak sınıflandırdığı örnekler (pozitif olarak tahmin edilen ama gerçekte negatif olan örnekler)
3. **True Negative (TN):** Modelin doğru bir şekilde negatif olarak sınıflandırdığı örnekler.
4. **False Negative (FN):** Modelin yanlış bir şekilde negatif olarak sınıflandırdığı örnekler (negatif olarak tahmin edilen ama gerçekte pozitif olan örnekler)

Araştırmada ikili sınıflandırma problemlerinde, modelin pozitif sınıfı ne kadar iyi ayırt edebildiğini değerlendirebilmek için ROC (Receiver Operating Characteristic) eğrisi kullanılmıştır. ROC eğrisi, farklı eşiklerdeki performansı görmeye imkan tanımaktadır.Modellerin her birinin sınıflandırma raporu oluşturularak performans değerleri gösterilmiştir. Sınıflandırma raporu, genellikle her bir sınıf için hassasiyet, geri çağırma ve F1 skoru gibi değerleri gösterir. Bu değerler, modelin her bir sınıfı ne kadar doğru ve kapsamlı bir şekilde tahmin ettiğini gösterir. Bu rapor, modelin performansını daha ayrıntılı bir şekilde anlaşılmasına yardımcı olur ve sınıflandırma problemiyle ilgili değerli içgörüler sağlar.

4.2.1 Lojistik Regresyon(Logistic Regression)

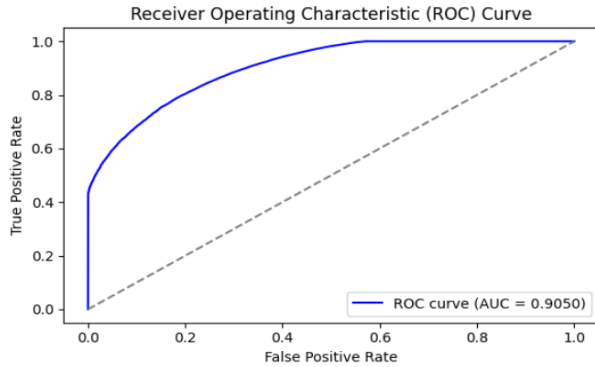
Scikit-learn kütüphanesinden LogisticRegression (lojistik regresyon modeli) ,confusion_matrix (karışıklık matrisi) ve classification_report(sınıflandırma raporu) temin edilerek kullanılmıştır.Seaborn ve matplotlib kütüphaneleri, veri görselleştirme amacıyla kullanılmıştır. LogisticRegression(solver='liblinear') fonksiyonu ile lojistik regresyon modeli oluşturulmuştur. Determinasyon katsayısı (R² skoru) hesaplanarak modelin eğitim seti üzerindeki performansı değerlendirilmiştir. Test verisi üzerinden tahminler yapılmış ve tahmin edilen değerlerin dağılımı analiz edilmiştir. Gerçek değerler ile tahmin edilen değerlerin karşılaştırılması için confusion matrix oluşturulmuştur.



Şekil 12.Logistik Regresyon Matrisi(Logistic Regression Matrix)

Lojistik Regresyon modelinin matrisdeki değerleri incelendiğinde 534 yanlış negatif (False Negative) tahmin yapıldığını ve oldukça düşük bir değer olduğu için modelin genellikle kredi ödememe ihtimali olanları doğru tespit edebildiğini göstermektedir.

ROC eğrisi ve AUC (Area Under Curve) skoru hesaplanmıştır. Predict_proba() fonksiyonu ile olası tahminlerin pozitif sınıfa ait olasılıkları elde edilmiştir. AUC skoru, modelin tahmin doğruluğunu özetleyen önemli bir metriktir ve 1'e ne kadar yakınsa model o kadar başarılıdır.



Şekil 13.Logistik Regresyon ROC Eğrisi (Logistic Regression ROC Curve)

Modelin doğruluk (accuracy) skoru 0.89, Auc-Roc skoru 0.91 olup bu performans sonuçlarıyla güçlü bir tahmin yeteneğine sahip olduğunu göstermektedir.

Sınıflandırma raporu ile modelin hassasiyet (precision), duyarlılık (recall), F1 skoru gibi temel metrikleri hesaplanmıştır. Sınıf 0 için precision yüksek (0.96) ancak recall düşük (0.45) olarak gözlenmiştir. Bu, modelin gecikmesi olmayanları doğru tahmin etme eğiliminde olduğunu ancak bunları tam olarak yakalayamadığını gösteriyor. Sınıf 1 için oldukça iyi performans göstermiştir (f1-score = 0.94). Bu model, gecikmesi olanları iyi tahmin ederken gecikmesi olmayanları kaçırıyor olarak değerlendirilebilmektedir.

Classification Report:

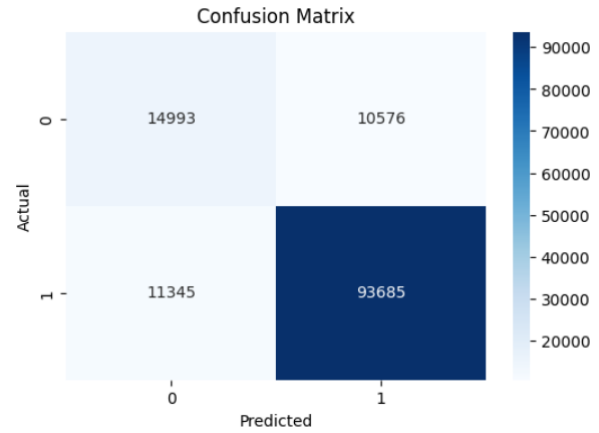
	precision	recall	f1-score	support
0	0.96	0.45	0.62	25569
1	0.88	0.99	0.94	105030
accuracy			0.89	130599
macro avg	0.92	0.72	0.78	130599
weighted avg	0.90	0.89	0.87	130599

Şekil 14.Logistik Regresyon Modeli Sınıflandırma Raporu(Logistic Regression Model Classification Report)

4.2.2 Karar Ağacı (Decision Tree)

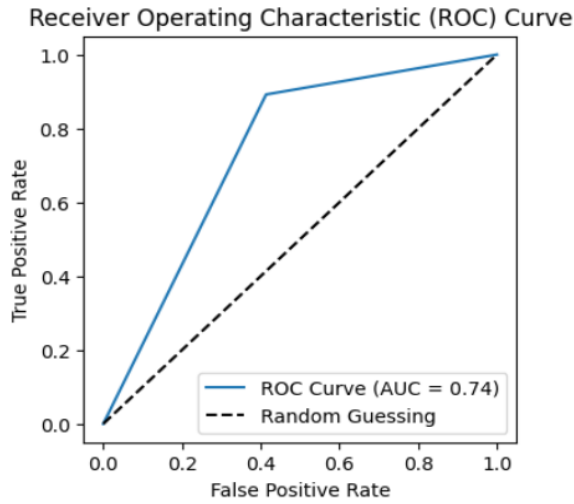
Karar Ağacı (Decision Tree) algoritması kullanılarak bir sınıflandırma modeli oluşturulmuş ve performans değerlendirilmesi yapılmıştır. Model, ID3 (Entropy) kriteri ile eğitilmiş ve test verileri üzerinde tahminleme gerçekleştirilmiştir. Scikit-learn (sklearn) karar ağacı modeli (DecisionTreeClassifier) ve model başarımları kullanılmıştır.

Karar ağacı modelinin matrisdeki değerleri incelendiğinde pozitif sınıfları tespit etmede yani düşük False Negatif ve yüksek False Pozitif tahminlerinde Logistic regresyon modeline kıyasla daha başarılı olduğu gözlenmiştir. Ancak negatif sınıfları yanlış pozitif olarak tahmin etme eğilimi daha yüksektir.



Şekil 15.Karar Ağacı Matrisi(Decision Tree Matrix)

Karar ağacı modelinin performans sonucuna göre accuracy skoru 0.83 ,ROC Auc Score değeri 0.74 'dür.Logistik regresyon modeline kıyasla daha düşük performans sonuçları elde edilmiştir.



Şekil 16. Karar Ağacı ROC Eğrisi (Decision Tree ROC Curve)

Sınıflandırma raporuna göre sınıf 0 için precision ve recall dengeli ama düşük. Sınıf 1 için f1-score biraz düşük olsa da (0.90), recall biraz daha dengeli görülmektedir. Gecikmesi olmayanları Logistic Regression'a kıyasla biraz daha iyi yakalıyor, ancak genel doğruluk açısından Logistic Regression kadar güçlü görülmemiştir.

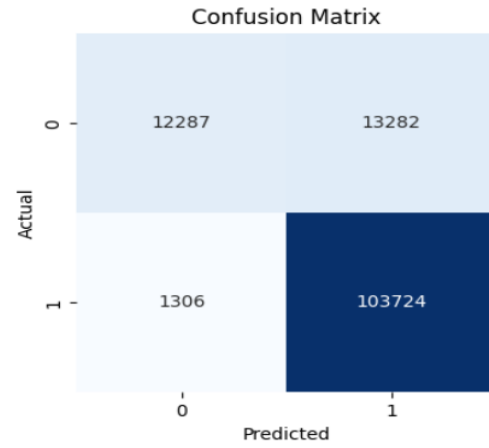
Classification Report:				
	precision	recall	f1-score	support
0	0.57	0.59	0.58	25569
1	0.90	0.89	0.90	105030
accuracy			0.83	130599
macro avg	0.73	0.74	0.74	130599
weighted avg	0.83	0.83	0.83	130599

Şekil 17. Karar Ağacı Modeli Sınıflandırma Raporu (Decision Tree Model Classification Report)

4.2.3 Catboost Algoritması (CatBoost Algorithm)

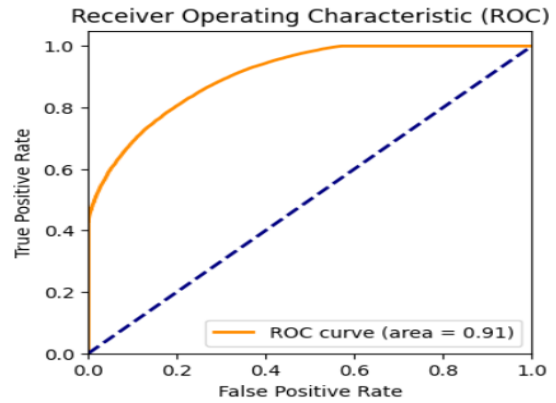
CatBoost (Categorical Boosting) algoritması kullanılarak bir sınıflandırma modeli oluşturulmuş ve performans değerlendirmesi gerçekleştirilmiştir. CatBoostClassifier, kategorik verilerle etkili çalışabilen bir gradient boosting algoritması olup, model eğitimi için varsayılan hiperparametreler kullanılmıştır.

Catboost modelinin matrisdeki değerleri incelendiğinde model, temerrüde düşen müşterileri büyük oranda doğru tahmin etmiş. Gerçekten temerrüde düşen müşterilerden sadece 1306'sını yanlış tahmin etmiş olarak görülmektedir.



Şekil 18. Catboost Matrisi (CatBoost Matrix)

Catboost modelinin performans sonucuna göre accuracy skoru 0.89 ROC Auc performans değeri 0.91 olup modelin güçlü bir tahmin yeteneğine sahip olduğunu gösterir.



Şekil 19. Catboost ROC Eğrisi (CatBoost ROC Curve)

Model genel olarak iyi bir eğitim ve test doğruluğuna sahip. Eğitim Doğruluğu 0.90, Test doğruluğu ise 0.89'dür. Ancak, test ROC Auc değeri, eğitim ROC Auc değerinden biraz daha düşüktür. Çıkan sonuçlar CatBoost'un test verisi üzerinde başarılı tahminler yaptığını ve geleneksel karar ağaçlarına kıyasla daha güçlü bir performans sunduğunu ortaya koymaktadır.

Test Classification Report:				
	precision	recall	f1-score	support
0	0.90	0.48	0.63	25569
1	0.89	0.99	0.93	105030
accuracy			0.89	130599
macro avg	0.90	0.73	0.78	130599
weighted avg	0.89	0.89	0.87	130599

Şekil 20. Catboost Modeli Sınıflandırma Raporu (Catboost Model Classification Report)

Sınıflandırma raporuna göre Sınıf 1'i Logistic Regression kadar iyi tahmin etmektedir. (F1-score = 0.93). Sınıf 0 için recall daha iyi (0.48), ancak precision biraz daha düşük olduğu görülmüştür.

4.2.4 Yapay Sinir Ağı (Artificial Neural Network)

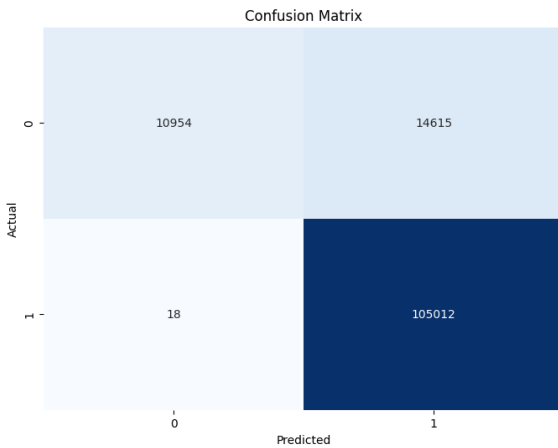
Çalışmada, Yapay Sinir Ağı (YSA) kullanılarak çok sınıflı bir sınıflandırma modeli oluşturulmuş ve değerlendirilmiştir. Model, Keras kütüphanesi ile Sequential API kullanılarak tasarlanmış ve Tam Bağlantılı Katmanlar (Fully Connected Layers) ile yapılandırılmıştır. Keras Yapay sinir ağı modeli (Sequential), tam bağlantılı katmanlar (Dense), aşırı öğrenmeyi önlemek için bırakma katmanı (Dropout) kullanılmıştır.

Modelin yapısı şu şekilde oluşturulmuştur:

- Giriş katmanı: 64 nöron, ReLU aktivasyon fonksiyonu.
- Gizli katmanlar: 64 nöronlu iki katman, ReLU aktivasyonu ve Dropout (0.5) kullanılarak aşırı öğrenme önlenmiştir.
- Çıkış katmanı: 3 sınıf için softmax aktivasyon fonksiyonu.

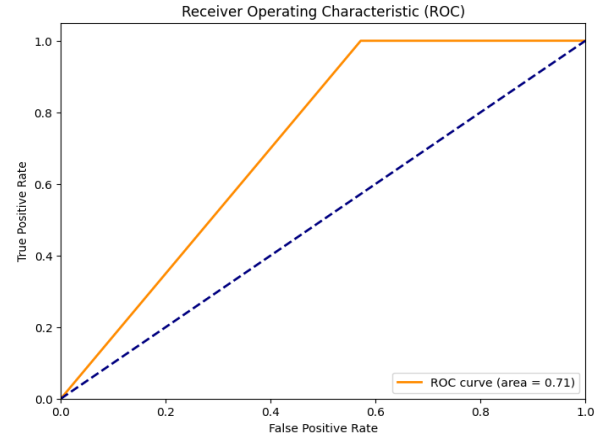
Model, Adam optimizasyon algoritması ve sparse_categorical_crossentropy kayıp fonksiyonu ile eğitilmiştir. 50 epoch ve 8 batch size kullanılarak eğitim gerçekleştirilmiş, doğrulama seti ile performans izlenmiştir.

Matrix 'e göre 10954 doğru negatif (True Negative), 14615 yanlış pozitif (False Positive), 18 yanlış negatif (False Negative) ve 105012 doğru pozitif (True Positive) tahmin yapıldığı görülmektedir. Model, borcunu ödeyemeyecek bireyleri (pozitif sınıf) çok iyi tespit ediyor (FN çok düşük, TP çok yüksek). Yanlış pozitiflerin (FP) fazla olması, bazı iyi müşterilere kredi vermeme riskini yaratabilir.



Şekil 21.Yapay Sinir Ağı Matrisi (Artificial Neural Network Matrix)

Yapay sinir ağı modelinin performans sonucuna göre Accuracy 0.89 ,ROC Auc Score değeri 0.71 'dir.



Şekil 22.Yapay Sinir Ağı ROC Eğrisi (Artificial Neural Network ROC Curve)

0 sınıfı için duyarlılık 0.43 ve 1 sınıfı için duyarlılık 1.00 olarak hesaplanmıştır. Bu, gerçek 0 sınıfına ait olan örneklerin yarısından azının doğru bir şekilde tahmin edildiğini ve gerçek 1 sınıfına ait olan örneklerin büyük bir kısmının doğru bir şekilde tahmin edildiğini gösterir. 1 sınıfı için yüksek bir F1-skoru elde edilmiştir.

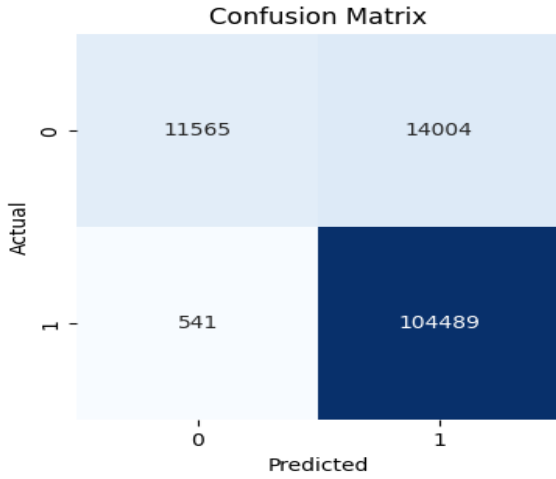
CLASSIFICATION REPORT:

	0	1	accuracy	macro avg	weighted avg
precision	1.00	0.88	0.89	0.94	0.90
recall	0.43	1.00	0.89	0.71	0.89
f1-score	0.60	0.93	0.89	0.77	0.87
support	25569.00	105030.00	0.89	130599.00	130599.00

Şekil 23.Yapay Sinir Ağları Modeli Sınıflandırma Raporu(Artificial Neural Networks Model Classification Report)

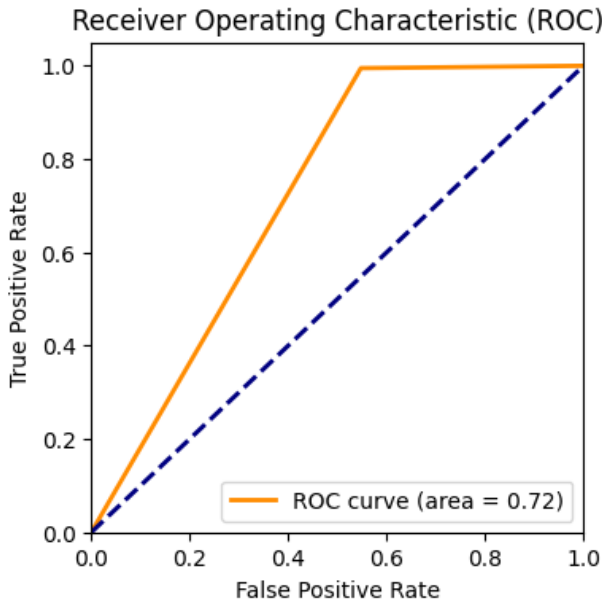
4.2.5 Random Forest Algoritması(Random Algoritması)

Rastgele Orman (Random Forest) algoritması kullanılarak bir sınıflandırma modeli oluşturulmuş ve performansı değerlendirilmiştir. RandomForestClassifier yöntemi, 100 karar ağacından (n_estimators=100) oluşan bir ansamble modeli olarak yapılandırılmıştır.Scikit-learn (sklearn): Random Forest modeli (RandomForestClassifier), modelin performans metrikleri (confusion_matrix, roc_curve, roc_auc_score, ConfusionMatrixDisplay) kullanılmıştır.



Şekil 24. Random Forest Matrisi (Random Forest Matrix)

Model temerrüde düşecek müşterileri (TP) büyük oranda doğru tahmin etmektedir. FN düşük olduğu için, riskli müşterileri kaçırma olasılığı düşük. Yanlış pozitif (FP) yüksek, yani bazı iyi müşterilere kredi verilmeme riski bulunmaktadır.



Şekil 25. Random Forest ROC Eğrisi (Random Forest ROC Curve)

Random Forest modelinin performans sonucuna göre Accuracy 0.89, ROC Auc Score değeri 0.72 'dir.

CLASSIFICATION REPORT:

	0	1	accuracy	macro avg	weighted avg
precision	0.96	0.88	0.89	0.92	0.90
recall	0.45	0.99	0.89	0.72	0.89
f1-score	0.61	0.93	0.89	0.77	0.87
support	25569.00	105030.00	0.89	130599.00	130599.00

Şekil 26. Random Forest Modeli Sınıflandırma Raporu (Random Forest Model Classification Report)

Sınıflandırma raporu incelendiğinde model eğitim verileri üzerinde çok yüksek bir doğruluk gösteriyor ancak test verileri üzerindeki doğruluk biraz daha düşük değerlerde görünüyor. Bu durum, modelin eğitim verilerine aşırı uyum sağlamış olabileceğini veya genelleme yeteneğinin biraz zayıf olabileceğini gösterebilir. Test sonuçlarına göre 0 sınıfı için kesinlik 0.95, 1 sınıfı için kesinlik 0.88 olarak hesaplanmıştır. Modelin 1 olarak tahmin ettiği örneklerin çoğunluğunun gerçekten 1 olduğunu ve 0 olarak tahmin ettiği örneklerin büyük bir kısmının da gerçekten 0 olduğunu gösterir. 0 sınıfı için duyarlılık 0.45 ve 1 sınıfı için duyarlılık 0.99 olarak hesaplanmıştır

5. SONUÇ VE ÖNERİLER (CONCLUSION AND RECOMMENDATIONS)

Veri analizi sonucunda seçilen öznitelik değişkenleri kullanılarak, kredilerin gecikme riski bir nevi kredi riski (0-1) bahsi geçen sınıflandırma algoritmaları ile tespit edilmeye çalışılmıştır. Kullanılan algoritmaların sınıflandırma performansları aşağıdaki tabloda belirtilerek karşılaştırılmıştır. Elde edilen nihai değişkenler ile eğitim, doğrulama ve test veri setleri üzerinden algoritmalar kredinin tahsis edilebilmesi için ,kredinin ödenebilme performansı yönünden müşterileri iyi ve kötü müşteri olarak sınıflandırarak, karmaşıklık matrisleri elde edilmiştir. Test veri seti üzerinden her bir algoritmaya ait kredi temerrüt (1) sınıfına ait karmaşıklık matrisinden elde edilen FN ,FP ve sınıflandırma raporundan elde edilen sınıflandırma ölçülerinin sonuçlarına aşağıdaki tabloda yer verilmiştir.

Model	Accuracy (%)	AUC-ROC	Precision	Recall	F1-Score	FN	FP
Lojistik Regresyon	89	0.91	0.88	0.99	0.94	534	13,954
Decision Tree	83	0.74	0.90	0.89	0.90	11,345	10,576
CatBoost	89	0.91	0.89	0.99	0.93	18	14,615
Yapay Sinir Ağları	89	0.71	0.88	1.00	0.93	18	14,615
Random Forest	89	0.72	0.88	0.99	0.93	541	14,004

Tablo 1 : Sınıflandırma Ölçütleri (Classification Metrics)

AUC-ROC (Receiver Operating Characteristic - Area Under Curve), modelin sınıfları ne kadar iyi ayırdığını gösterir. Lojistik Regresyon ve CatBoost en yüksek AUC-ROC değerine sahip modellerdir. Recall (1), gerçekten temerrüde düşen müşterilerin ne kadarını doğru tahmin ettiğini gösterir. Yapay Sinir Ağları (1.00) recall açısından en iyi modeldir. Ancak, diğer metrikleri de değerlendirmek gerekir. Precision (1), modelin "temerrüde düşecek" dediği kişilerin gerçekten temerrüde düşüp düşmediğini gösterir. Decision Tree bu metrikte en yüksek değere sahip olsa da genel başarısı düşüktür. FN, modelin temerrüde düşeceğini öngöremediği müşteri sayısını gösterir. Düşük olması istenir. En düşük FN değerine sahip modeller CatBoost ve Yapay Sinir Ağlarıdır. Genel olarak çalışma sonucu değerlendirildiğinde CatBoost modeli, kredi temerrüdü tahmini için en iyi seçenek olduğu söylenebilir.

☑ **Yüksek AUC-ROC (0.91):** Modelin temerrüde düşenleri ve düşmeyenleri ayırma gücü yüksektir.

☑ **Recall (1) = 0.99:** Temerrüde düşecek kişileri kaçırma oranı çok düşük.

☑ **FN = 18 (En düşük değerlerden biri):** Yanlış negatif

sayısı minimum seviyede.
☒ **Denekli Precision ve F1-Score değerleri:** Overfitting yapmadan doğru tahmin yaptığı olarak değerlendirilmektedir.

Yapılan çalışma neticesinde hem yüksek AUC-ROC skoruna sahip hem de en düşük FN değerine sahip CatBoost modelinin kredi temerrüdünü tahmin etmek için en uygun model olduğu değerlendirilmiştir. Çalışma geleneksel lojistik regresyon yerine CatBoost gibi güçlü modeller de kullanılarak literatürdeki klasik yöntemlere kıyasla başarılı performanslar elde edilebileceğini göstermesi açısından önemlidir.

KAYNAKLAR (REFERENCES)

- [1] Internet: C.Yeh & C.H Lien, The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients, <https://www.sciencedirect.com/science/article/abs/pii/S0957417407006719>, 11.11.2024.
- [2] Internet: H. Budak & S. Erpolat, Kredi Riski Tahmininde Yapay Sinir Ağları ve Lojistik Regresyon Analizi Karşılaştırılması., <https://dergipark.org.tr/tr/pub/ajit-e/issue/54448/741024>, 01.10.2024.
- [3] Internet: Ş. Kavcıoğlu, Ticari Bankacılıkta Kredi Riskinin ve Kredi Riski Ölçüm Modellerinin Değerlendirilmesi, <https://dergipark.org.tr/tr/download/article-file/3978>, 02.10.2024.
- [4] Internet: S. Hamori & M. Kawai & K. Takahiro & M. Yuji & C. Watanabe, Ensemble Learning or Deep Learning Application to Default Risk Analysis, https://epe.lacbac.gc.ca/100/201/300/jrm_risk_financial_mngmt/2018/v11no1/mdpi.com/1911-8074/11/1/12/htm.html, 11.10.2024.
- [5] Internet: T. İldaş, Kredi Riski Ölçüm Modellerinin Değerlendirilmesi, <https://dergipark.org.tr/en/download/article-file/1901150>, 15.10.2024.
- [6] M. Mahbobi & S. Kimiagari & M. Vasudevan, "Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks", *Annals of Operations Research*, 330(1), 609–637, 2023.
- [7] Internet: T.E Tütüncü & S. Gürsakal, Kredi Temerrüt Riskini Tahmin Etmede Makine Öğrenme Algoritmalarının Karşılaştırılması, <https://dergipark.org.tr/tr/download/article-file/2635086>, 01.11.2024.
- [8] Internet: M.E.F Milli, Sınıf Dengeleme Yöntemlerinin Makine Öğrenmesi Teknikleri Üzerine Etkisi: Kredi Risk Örneği, <https://avesis.deu.edu.tr/yonetilen-tez/85127e15-e6b5-4804-a3b5-d6eae4aa02f9/sinif-dengeleme-yontemlerinin-makine-ogrenmesi-teknikleri-uzerine-etkisi-kredi-risk-ornegi>, 05.10.2024.
- [9] Internet: X.Zhu & O. Chu & X.Song & P.Hu & L.Peng, Explainable prediction of loan default based on machine learning models, <https://www.sciencedirect.com/science/article/pii/S2666764923000218>, 11.10.2024.