



Makine Öğrenmesi Yöntemlerini Kullanarak Kredi Kartı Onay / Red Analizi

Yazılım Mühendisliği Anabilim Dalı

Dönem Projesi

Emre YAŞA

ORCID 0009-0001-3490-9306

Proje Danışmanı: Doç. Dr. Sıla Övgü KORKUT UYSAL

Ocak 2024

Makine Öğrenmesi Yöntemlerini Kullanarak Kredi Kartı Onay / Red Analizi

ÖZ

Ticari bankalara kredi kartları için çok sayıda başvuru gelmektedir. Birçoğu, örneğin yüksek kredi bakiyeleri, düşük gelir seviyeleri veya bir kişinin kredi raporuna ilişkin çok fazla sorgu yapılması gibi birçok nedenden dolayı reddedilmektedir.

Kaliforniya Üniversitesi Irvine Makine Öğrenimi Havuzundaki kredi kartı onay veri kümesi üzerinde onay/red sonucu için önemli olan faktörler belirlenerek, Naive Bayes, Destek Vektör Makineleri, Gradyan Arttırma, Karar Ağacı ve Rastgele Orman Sınıflandırıcısı gibi makine öğrenme algoritmaları denenecektir.

Uygulanan algoritmalar doğruluk, duyarlılık ve hassasiyet ölçütleri kıyaslanarak değerlendirilecektir. En yüksek başarıyı veren algoritma ile red ve onay durumları paylaşılacaktır.

Anahtar Sözcükler: Naive Bayes, Destek Vektör Makineleri, Gradyan Arttırma, Karar Ağacı Sınıflandırıcısı, Rastgele Orman Sınıflandırıcısı, Doğruluk, Duyarlılık, Hassasiyet, Kredi Kartı Onayı

Credit Card Approval / Rejection Analysis Using Machine Learning Methods

Abstract

Many applications are received for credit cards at commercial banks. Many of them are rejected for various reasons, such as high credit balances, low income levels, or excessive inquiries into a person's credit report.

By determining the factors that are crucial for approval/rejection outcomes on the credit card approval dataset in the University of California Irvine Machine Learning Repository, machine learning algorithms such as Naive Bayes, Support Vector Machines, Gradient Boosting, Decision Tree, and Random Forest Classifier will be experimented with.

The applied algorithms will be evaluated by comparing accuracy, recall, and precision metrics. The algorithm that yields the highest performance will be shared along with the cases of approval and rejection.

Keywords: Naive Bayes, Support Vector Machines, Gradient Boosting, Decision Tree Classifier, Random Forest Classifier, Accuracy, Recall, Precision, Credit Card Approval

Teşekkür

Bu çalışmanın gerçekleştirilmesi sürecinde değerli destekleri ve katkıları için saygıdeğer danışman hocam Doç. Dr. Sıla Övgü KORKUT UYSAL'a; 3 dönem boyunca değerli bilgilerini bizlerle paylaşan, sonsuz anlayışı ve yardımları için Sn. Prof. Dr. Aytuğ ONAN'a sonsuz teşekkürlerimi sunarım.

İçindekiler

Öz	i
Abstract	ii
Teşekkür	iii
Formüller Listesi.....	vi
Tablolar Listesi.....	vii
Kısaltmalar Listesi	viii
1 Giriş	1
1.1 Problem Tanımı.....	1
1.2 Literatür Taraması	1
1.3 Veri Kümesi.....	2
1.4 Yöntem	3
1.5 Amaç	4
1.6 Çalışmanın Katkıları	4
2 Kullanılan Algoritmalar.....	5
2.1 Naive Bayes	5
2.2 Destek Vektör Makineleri	6
2.3 Gradyan Arttırma	6
2.4 Karar Ağacı Sınıflandırıcısı	8
2.5 Rastgele Orman Sınıflandırıcısı	10
3 Kullanılan Ölçütler.....	11
3.1 Doğruluk.....	11
3.2 Hassasiyet	11
3.3 Duyarlılık.....	12
4 Bulgular	14

5	Model Performansı Deęerlendirmeleri.....	15
5.1	Naive Bayes	15
5.2	Gradyan Arttırma	15
5.3	Destek Vektör Makineleri	15
5.4	Karar Ağacı Sınıflandırıcısı	15
5.5	Rastgele Orman Sınıflandırıcısı	15
6	Sonuç.....	16
	Kaynaklar	17

Formüller Listesi

Formül 2.1: Bayes Teoremi.....	3
Formül 2.2: Lineer SVM.....	4
Formül 2.3: Gradyan Arttırma Zayıf Öğrenme Formülü	5
Formül 2.4: Gradyan Arttırma Genel Formülü	5
Formül 2.5: Karar Ağaçları Sınıflandırıcısı Formülü	7
Formül 2.6: Karar Ağaçları Sınıflandırıcısı Sınıf Etiketleri Tahmin Formülü.....	7
Formül 2.7: Rastegele Orman Sınıflandırıcısı Formülü	8
Formül 2.7: Rastegele Orman X verildiğinde C sınıfının hesaplanması	8
Formül 3.1: Doğruluk Ölçütü Formülü.....	9
Formül 3.2: Hassasiyet Ölçütü Formülü	10
Formül 3.3: Duyarlılık Ölçütü Formülü.....	11

Tablolar Listesi

Tablo T.1: Ölçüt Değerleri	11
----------------------------------	----

Kısaltmalar Listesi

SVM	Destek Vektör Makineleri
FBE	Fen Bilimleri Enstitüsü
İKÇÜ	İzmir Kâtip Çelebi Üniversitesi
NB	Naive Bayes
ORCID	Open Researcher and Contributor ID
UCI	Kaliforniya Üniversitesi Irvine

Bölüm 1

Giriş

1.1 Problem Tanımı

Kredi kartı, günümüzde en yaygın kullanılan ödeme yöntemlerinden biridir. Kredi kartlarının kullanımının artması ile birlikte, kredi kartı başvurularının değerlendirilmesi de önemli bir hale gelmiştir. Kredi kartı başvurularının değerlendirilmesinde, başvuru sahibinin geliri, kredi puanı ve ödeme geçmişi gibi faktörler göz önünde bulundurulur. Bu faktörleri dikkate alarak, başvurunun onaylanması veya reddedilmesine karar verilir. Bu süreç, hem müşteri memnuniyetini artırmak hem de finansal riskleri minimize etmek amacıyla etkin bir şekilde yönetilmelidir. Bu çalışma, kredi kartı başvurularının onaylanma durumunu öngörmek için makine öğrenimi tekniklerini kullanmayı amaçlamaktadır.

1.2 Literatür Taraması

Yapılan literatür taramasında, çalışmalarda genel olarak kredi risk tahminleri yapılmış, bu çalışmadaki gibi bankanın müşterinin başvurusunu kabul etme tahminlemesi yapılmıştır.

[2] Arun, Ishaan ve Sanmeet (2016) tarafından yapılan çalışmalarda müşterilerin başvuruları değerlendirilerek krediye en uygun müşterinin tahminlemesi amaçlanmış ve çalışmalar yapılmıştır. Cinsiyet, medeni durum, vasilik, eğitim durumu, gelir, ek gelir, başvuru kredi miktarı, kredi süresi, sahip olunan taşınmaz mallar gibi veriler üzerinden kredinin kabul edilip edilmeyeceğinin tahmini amaçlanmıştır. Metot olarak Karar Ağaçları, Rastgele Orman, Lineer Modeller ve Gaussian NB kullanılmış, sınıflandırma gerçekleştirilmiştir. Performans test sonuçları makalede paylaşılmayıp en iyi sonuç veren modelin Gaussian NB olduğu belirtilmiştir.

[3] Serengil, İmece, Tosun, Büyükbaş ve Köroğlu (2021) tarafından yapılan çalışmada tahsili geciken kredinin farklı sınıflandırma algoritmaları ile tahmin edilmesi ve karşılaştırılması amaçlanmıştır. Özel bir bankanın 181.276 adet müşteri verisinden oluşan bir veri seti kullanılmıştır. Veri setinde müşteri ödeme geçmişi, bilançolar, önceki kredi kartı ödemeleri, diğer bankalardan gelen risk ve limit verileri gibi özellikler bulunmaktadır. Sonuç olarak elde edilecek değerler sağlıklı, gecikmeli ve tahsili gecikmiş kredi değerleridir. Veri seti farklı modellerde test edilmiş ve karşılaştırmalar sonucunda problem için en uygun modelin 0.90 özgüllük, 0.87 kesinlik, 0.77 duyarlılık ve 0.82 F1 skoru olarak elde edilerek Karar Ağaçları algortiması olduğu belirtilmiştir.

[4] Sarızeybek ve Seveli (2022) tarafından yapılan çalışmada banka müşterilerinin kredi alma eğiliminin karşılaştırmalı analizi yapılmıştır. Doğruluk değeri en yüksek olan sınıflandırma modelinin %98,86 ile Rastgele Orman algoritması olduğu belirtilmiştir.

[5] Alan (2020) tarafından yapılan çalışmada, en iyi sınıflandırma modelinin belirlenmesi için çapraz doğrulama yöntemlerinden hold-out ve k-katlı çapraz doğrulama kullanılmış, bu işlemler 32 ayrı veri setine uygulanmıştır. Veri setlerinin arasında Portekiz Bankası Pazarlama veri setinde çapraz doğrulama ile en iyi modelin %88,75 doğruluk oranına sahip olan SVM sınıflandırıcısı olduğu tespit edilmiş, AUC oranı %53,32 ve F1 skor değeri ise %61,07 olarak elde edilmiştir. Veri seti üzerindeki çalışma sonucunda doğruluk değeri %86,35, AUC değeri %60,23 ve F1 skoru ise %61,83 olarak elde edilmiştir.

1.3 Veri Kümesi

Bu projede, Kaliforniya Üniversitesi Irvine Makine Öğrenimi Havuzu[1]'ndan elde edilen bir kredi kartı başvuru veri kümesi kullanılmıştır. Veri kümesi 690 satır ve 16 öznitelik barındırmaktadır. Bu öznitelikler sırasıyla, Cinsiyet, Yaş, Borç, Medeni Durum, Banka Müşterisi, Eğitim Seviyesi, Etnik Köken, İş Tecrübesi, Önceki Temerrüt, İş Durumu, Kredi Puanı, Ehliyet, Uyruk, Posta Kodu, Gelir ve Onay Durumu'dur. Ana odak noktası, başvuruların onay veya red edilme durumunu belirten

"Onay Durumu" özniteliğidir. Model performansını değerlendirmek için kullanılacak olan öznitelikler "Gelir" ve "Onay Durumu" dur.

1.4 Yöntem

İlk olarak veri seti yüklenip görüntülenmiştir. Veri kümesinin hem sayısal hem de sayısal olmayan özniteliklerin bir karışımına sahip olduğu ve ayrıca bir takım eksik girişler içerdiği tespit edilmiştir.

Veri setinin info bilgisi aşağıdaki gibidir.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 690 entries, 0 to 689
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Gender                690 non-null   object
1   Age                   690 non-null   object
2   Debt                  690 non-null   float64
3   Married               690 non-null   object
4   BankCustomer         690 non-null   object
5   EducationLevel       690 non-null   object
6   Ethnicity             690 non-null   object
7   YearsEmployed        690 non-null   float64
8   PriorDefault         690 non-null   object
9   Employed              690 non-null   object
10  CreditScore           690 non-null   int64
11  DriversLicense        690 non-null   object
12  Citizen               690 non-null   object
13  ZipCode               690 non-null   object
14  Income                690 non-null   int64
15  ApprovalStatus       690 non-null   object
dtypes: float64(2), int64(2), object(12)
memory usage: 86.4+ KB
None
```

Makine öğrenimi algoritmalarının iyi tahminler yapabilmesini sağlamak için veri kümesi üzerinde kolon bazlı farklı ön işleme stratejileri izlenmiştir. Örnek olarak, posta kodu kolonu tamamen kaldırılmıştır. Yaş kolonundaki eskik veriler cinsiyet bazlı ortalama alınarak, medeni durum kolonundaki eksik veriler “bekar” olarak, banka müşterisi kolonundaki eksik veriler ise “false” olarak doldurulmuştur.

Ön işleme sonrasında veri kümesi, eğitim ve test olarak %80 - %20 oranında ikiye ayrılmıştır. Eğitim kümesi, algoritmaların eğitilmesi için kullanılmıştır. Test kümesi ise algoritmaların performansının değerlendirilmesi için kullanılmıştır.

1.5 Amaç

Bu çalışmanın temel amacı, farklı makine öğrenimi algoritmalarını kullanarak kredi kartı başvurularının onay durumunu tahmin etmektir. Proje, ele alınan algoritmaların performansını karşılaştırarak en etkili olanını belirlemeyi hedeflemektedir.

1.6 Çalışmanın Katkıları

Bu çalışmanın ana katkıları şunlardır:

- Farklı makine öğrenimi algoritmalarının kredi kartı başvurularını onaylama yeteneğini karşılaştırmak.
- Gelir seviyesi, eğitim düzeyi gibi belirlenen özelliklerin başvuruların değerlendirilmesindeki etkisini incelemek.

Bölüm 2

Kullanılan Algoritmalar

Bu çalışmada, Naive Bayes, Destek Vektör Makineleri, Gradyan Arttırma, Karar Ağacı ve Rastgele Orman Sınıflandırıcısı gibi makine öğrenme algoritmaları kullanılmıştır.

2.1 Naive Bayes

Naive Bayes algoritması, Bayes teoremini kullanarak çalışır. Bayes teoremine göre, bir olayın olasılığı, olayın gerçekleşme olasılığı ile olayın gerçekleşmesi durumunda diğer olayın gerçekleşme olasılığının çarpımına eşittir.

Naive Bayes algoritması, Bayes teoremini kullanarak, her bir sınıf için olasılıkları hesaplar. Bu olasılıklar, sınıflandırma için kullanılır. Basit ve hızlı bir algoritmadır. Bu nedenle, büyük veri kümeleri üzerinde bile etkili bir şekilde kullanılabilir.

Naive Bayes algoritmasının temelleri aşağıdaki şekilde gösterilebilir:

$$P(\text{sınıf} | \text{özellikler}) = P(\text{özellikler} | \text{sınıf}) * P(\text{sınıf}) / P(\text{özellikler})$$

Formül 2.1: Bayes Teoremi

Bu denklemde:

- $P(\text{sınıf} | \text{özellikler})$: Sınıfın özellikler verildiğinde olasılığı
- $P(\text{özellikler} | \text{sınıf})$: Özelliklerin sınıf verildiğinde olasılığı
- $P(\text{sınıf})$: Sınıfın olasılığı
- $P(\text{özellikler})$: Özelliklerin olasılığı

Naive Bayes algoritması, özelliklerin bağımsız olduğu varsayımını yapar. Bu varsayım, gerçekte her zaman doğru olmayabilir. Ancak, bu varsayım oldukça iyi bir yaklaşımdır.

2.2 Destek Vektör Makineleri

Destek Vektör Makineleri (SVM), gözetimli öğrenme yöntemlerinden biridir. Sınıflandırma, regresyon ve kümeleme gibi çeşitli makine öğrenimi problemlerini çözmek için kullanılabilir. SVM'ler, bir doğru veya hiper düzlem kullanarak veri noktalarını sınıflandırmak için tasarlanmıştır. Bu doğru veya hiper düzlem, iki sınıfın noktaları için de maksimum uzaklıkta olacak şekilde seçilir.

SVM'ler, sınıflandırma problemlerinde genellikle en iyi performans gösteren yöntemlerden biridir. Bu, özellikle sınıflar iyi ayrılmışsa geçerlidir. SVM'ler ayrıca küçük veya orta büyüklükteki veri setleri için uygundur.

Lineer SVM'nin temel formülü, giriş verisi X ve W , b düzeltme terimi ve ağırlıklar olmak üzere şu şekildedir:

$$f(X) = W \cdot X + b$$

Formül 2.2: Lineer SVM

SVM, sınıflar arasındaki marjı maksimize etmeye çalışırken, bu formülü kullanarak bir optimizasyon problemi çözer.

2.3 Gradyan Arttırma

Gradyan Arttırma, denetimli öğrenmede kullanılan bir makine öğrenmesi algoritmasıdır. Bu algoritma, zayıf öğrenme modellerini birleştirerek güçlü bir

öğrenme modeli oluşturmayı amaçlar. Zayıf öğrenme modelleri, tahminleri genellikle doğru olan ancak tahmin güçleri sınırlı olan modellerdir.

Gradyan Arttırma algoritması, iteratif olarak çalışır. Her iterasyonda, yeni bir zayıf öğrenme modeli eklenir ve bu model, önceki modellerin hatalarını düzeltmeye çalışır. Bu süreç, tüm veriler üzerinde en iyi tahmini elde edene kadar devam eder.

Gradyan Arttırma algoritması, makine öğrenimi alanında yaygın olarak kullanılan güçlü bir algoritmadır. Regresyon ve sınıflandırma problemlerinde etkili bir şekilde kullanılabilir.

Gradyan Arttırma algoritmasında, her iterasyonda, yeni bir zayıf öğrenme modeli, kayıp fonksiyonunun gradyanını en aza indirecek şekilde eğitilir. Bu gradyan, aşağıdaki şekilde hesaplanır:

$$g(x_i) = -(y_i - h(x_i))$$

Formül 2.3: Gradyan Arttırma Zayıf Öğrenme Formülü

Bu denklemden:

- $g(x_i)$: x_i noktasındaki gradyanı temsil eder.
- y_i : i . gözlemin hedef değeri temsil eder.
- $h(x_i)$: x_i noktasındaki tahmini temsil eder.

Gradyan Arttırma algoritmasının genel formülü aşağıdaki şekildedir:

$$h(x) = h_0(x) + \sum_{i=0}^T g(x_i)$$

Formül 2.4: Gradyan Arttırma Genel Formülü

Bu denklemde:

- $h(x)$: x noktasındaki tahmini temsil eder.
- $h_0(x)$: ilk zayıf öğrenme modelinin tahmini temsil eder.
- $g(x_i)$: i . zayıf öğrenme modelinin gradyanı temsil eder.
- T : toplam zayıf öğrenme modeli sayısını temsil eder.

2.4 Karar Ağacı Sınıflandırıcısı

Karar ağacı sınıflandırıcısı, makine öğrenimi alanında yaygın olarak kullanılan bir sınıflandırma algoritmasıdır. Bu algoritma, verilerdeki kalıpları belirlemek için bir ağaç yapısı kullanır.

Karar ağacı sınıflandırıcısı, aşağıdaki adımları izleyerek çalışır:

1. İlk olarak, veri seti, her biri bir sınıf etiketi ile etiketlenmiş kayıtlardan oluşan bir dizi olarak temsil edilir.
2. Ardından, ağaç yapısının kökü, veri setini en iyi şekilde bölen özellik seçilerek oluşturulur.
3. Kökten gelen her dal, kök özelliğinin belirli bir değerine sahip kayıtları içerir.
4. Bu işlem, her dalın artık bölünemez hale gelene kadar tekrarlanır.

Karar ağacı sınıflandırıcısı, yeni bir kayıt için sınıf etiketini tahmin etmek için aşağıdaki adımları izler:

1. Kayıt, ağacın kökünden başlar.
2. Kayıttaki değer, ağacın kök özelliğinin değerine göre karşılaştırılır.
3. Kayıt, karşılaştırma sonucuna göre ağacın ilgili dalına gider.
4. Bu işlem, kayıt bir yaprak düğüme ulaşana kadar tekrarlanır.

Karar ağacı sınıflandırıcısının formülasyonu, ağacın yapısı ve sınıf etiketlerini tahmin etmek için kullanılan kurallarla ilgilidir.

Ağacın yapısı, aşağıdaki formül ile tanımlanabilir:

$$T = (D, L, R, E)$$

Formül 2.5: Karar Ağaçları Sınıflandırıcısı Formülü

Bu denklemde:

- D: Veri seti
- L: Kök düğüm
- R: Dallar
- E: Yaprak düğümler

Kök düğüm, veri setini en iyi şekilde bölen özelliği temsil eder. Dallar, kök özelliğinin belirli bir değerine sahip kayıtları içerir. Yaprak düğümler, kayıtlar için tahmin edilen sınıf etiketlerini içerir.

Sınıf etiketlerini tahmin etmek için kullanılan kurallar, aşağıdaki formül ile tanımlanabilir:

$$y = f(x, T)$$

Formül 2.6: Karar Ağaçları Sınıflandırıcısı Sınıf Etiketleri Tahmin Formülü

Bu denklemde:

- y: Tahmin edilen sınıf etiketi
- x: Kayıt
- T: Karar ağacı

Bu formülde, f fonksiyonu, ağacın yapısı ve sınıf etiketlerini tahmin etmek için kullanılan kuralları temsil eder.

2.5 Rastgele Orman Sınıflandırıcısı

Rastgele orman sınıflandırıcısı, bir veri kümesi üzerinde bir sınıflandırma görevi gerçekleştirmek için kullanılan bir makine öğrenmesi algoritmasıdır. Bu algoritma, bir dizi karar ağacı oluşturur ve ardından bu ağaçların kararlarını birleştirerek bir sınıflandırma tahmini yapar.

Rastgele orman sınıflandırıcısının formülasyonu şu şekildedir:

$$y = \operatorname{argmax}_c P(c|x)$$

Formül 2.7: Rastgele Orman Sınıflandırıcısı Formülü

Bu denklemde:

- y , sınıflandırma tahminidir.
- c , olası sınıflardır.
- $P(c|x)$, x verildiğinde c sınıfının olasılığıdır.

Rastgele orman sınıflandırıcısı, x verildiğinde c sınıfının olasılığını şu şekilde hesaplar:

$$P(c|x) = 1/K * \sum_{i=1}^K I(c = c_i)P(x|c_i)$$

Formül 2.8: Rastgele Orman Sınıflandırıcısı X verildiğinde C sınıfının hesaplanması

Bu denklemde:

- K , karar ağacı sayısıdır.
- c_i : i . karar ağacının sınıflandırıcısıdır.
- $I(c = c_i)$, c ve c_i 'nin eşit olup olmadığını belirten bir fonksiyondur.
- $P(x)$, x verildiğinde c_i sınıfının olasılığıdır.

Bölüm 3

Kullanılan Ölçütler

Bu çalışmada, doğruluk, hassasiyet ve duyarlılık ölçütleri tespit edilmiştir.

3.1 Doğruluk

Makine öğrenmesi doğruluk ölçütü, bir makine öğrenmesi modelinin performansını değerlendirmek için kullanılan bir ölçüttür. Bu ölçüt, modelin doğru tahmin ettiği örneklerin yüzdesini temsil eder.

Doğruluk ölçütü, sınıflandırma problemlerinde yaygın olarak kullanılan bir ölçüttür. Bu tür problemlerde, modele bir örnek verildiğinde, model bu örneği belirli bir sınıfa ait olup olmadığını tahmin eder.

Doğruluk ölçütü, aşağıdaki şekilde hesaplanır:

$$\text{Doğruluk} = (\text{Doğru Tahminler}) / (\text{Toplam Örnek Sayısı})$$

Formül 3.1: Doğruluk Ölçütü Formülü

Doğruluk ölçütü, bir makine öğrenmesi modelinin genel performansını değerlendirmek için basit ve kullanışlı bir ölçüttür.

3.2 Hassasiyet

Makine öğrenmesi hassasiyeti ölçütü, bir makine öğrenmesi modelinin performansını değerlendirmek için kullanılan bir ölçüttür. Bu ölçüt, modelin pozitif olarak tahmin etmesi gereken örnekleri doğru olarak tahmin etme yüzdesini temsil eder.

Duyarlılık, sınıflandırma problemlerinde yaygın olarak kullanılan bir ölçüttür. Bu tür problemlerde, modele bir örnek verildiğinde, model bu örneği belirli bir sınıfa ait olup olmadığını tahmin eder.

Hassasiyet ölçütü aşağıdaki şekilde hesaplanır:

$$\text{Hassasiyet} = (\text{Doğru Pozitif Tahminler}) / (\text{Toplam Pozitif Tahminler})$$

Formül 3.2: Hassasiyet Ölçütü Formülü

Hassasiyet ölçütü, bir makine öğrenmesi modelinin pozitif olarak tahmin ettiği örneklerin ne kadar doğru olduğunu değerlendirmek için kullanılır. Bu ölçüt, özellikle modelin negatif olarak tahmin ettiği örneklerin sayısını azaltmaya odaklanan uygulamalarda önemlidir.

3.3 Duyarlılık

Makine öğrenmesi duyarlılık ölçütü, bir makine öğrenmesi modelinin performansını değerlendirmek için kullanılan bir ölçüttür. Bu ölçüt, modelin pozitif olarak tahmin etmesi gereken örneklerin yüzdesini temsil eder.

Duyarlılık ölçütü, sınıflandırma problemlerinde yaygın olarak kullanılan bir ölçüttür. Bu tür problemlerde, modele bir örnek verildiğinde, model bu örneği belirli bir sınıfa ait olup olmadığını tahmin eder.

Duyarlılık ölçütü, aşağıdaki şekilde hesaplanır:

$$\text{Duyarlılık} = (\text{Doğru Pozitif Tahminler}) / (\text{Gerçek Pozitifler})$$

Formül 3.3: Duyarlılık Ölçütü Formülü

Duyarlılık ölçütü, bir makine öğrenmesi modelinin pozitif olarak tahmin etmesi gereken örneklerin ne kadar iyi tahmin ettiğini değerlendirmek için kullanılır. Bu ölçüt, özellikle modelin negatif olarak tahmin ettiği örneklerin sayısını azaltmaya odaklanan uygulamalarda önemlidir.

Bölüm 4

Bulgular

Bu çalışma, UCI Makine Öğrenimi Havuzundaki kredi kartı onay veri kümesi üzerinde farklı makine öğrenme algoritmalarının performansını değerlendirmeyi amaçlamıştır.

İncelenen beş farklı makine öğrenmesi modelinin ölçüt değerleri aşağıdaki gibidir:

Model	Doğruluk	Duyarlılık	Hassasiyet
Gaussian NB	0.647	0.196	0.786
Gradyan Arttırma	0.669	0.482	0.628
SVC	0.676	0.357	0.714
Karar Ağacı	0.676	0.5	0.636
Rastgele Orman	0.669	0.5	0.622

Tablo T.1: Ölçüt Değerleri

Bölüm 5

Model Performansı Değerlendirmeleri

Proje kapsamında kullanılan kredi kartı onay veri kümesi, toplamda 690 örnek içermektedir. Bu örneklerin 307 tanesi onaylanmış, 383 tanesi ise reddedilmiştir.

5.1 Gaussian NB

Gaussian NB modelinin doğruluk oranı makul, ancak duyarlılık ve hassasiyet düşük görünmektedir. Bu, modelin onayları düşük bir başarı oranıyla tespit ettiğini ve reddedilenleri sıklıkla yanlış sınıflandırdığını göstermektedir.

5.2 Gradyan Arttırma

Gradyan Arttırma modeli, genel olarak makul bir doğruluk oranı sunarken, duyarlılık ve hassasiyet oranları daha dengeli bir performans göstermektedir.

5.3 Destek Vektör Makineleri

Destek Vektör Makineleri modelinin olumsuz örnekleri daha iyi sınıflandırdığı, ancak olumlu örnekleri kaçırma eğiliminde olduğu gözlemlenmiştir.

5.4 Karar Ağacı Sınıflandırıcısı

Karar Ağacı sınıflandırıcısı, veri kümesini genel olarak iyi sınıflandırmış ve duyarlılık ile hassasiyet oranlarında dengeli bir performans sergilemiştir.

5.5 Rastgele Orman Sınıflandırıcısı

Rastgele Orman Sınıflandırıcısı, Karar Ağacı Sınıflandırıcısı modeli ile benzer bir performans sergilemiştir.

Bölüm 6

Sonuç

Bu proje kapsamında, kredi kartı onay veri kümesi üzerinde beş farklı makine öğrenimi algoritmasıyla sınıflandırma çalışmaları gerçekleştirilmiştir. Elde edilen sonuçlar, her bir modelin farklı avantajlara ve zorluklara sahip olduğunu göstermektedir.

GaussianNB modeli, doğruluk oranı bakımından makul bir performans sergilese de, onayları düşük bir doğrulukla tespit etmiş ve reddedilenleri sıklıkla yanlış sınıflandırmıştır. Diğer yandan, Gradyan Arttırma ve Karar Ağacı Sınıflandırıcısı modelleri özellikle duyarlılık ve hassasiyet oranları açısından daha dengeli bir performans göstermiştir.

SVC modeli, olumsuz örnekleri daha iyi sınıflandırmış, ancak onayları kaçırma eğiliminde olmuştur. Rastgele Orman Sınıflandırıcısı ise genel olarak diğer modellerle benzer bir performans sergilemiştir.

Sonuç olarak, proje kapsamında kullanılan veri seti ve algoritmaların kombinasyonlarına bağlı olarak farklı performanslar elde edilmiştir. Bu bulgular, projenin amacına ve önceliklerine bağlı olarak model seçimi yapılırken dikkate alınmalıdır. Gelecekteki çalışmalarda, daha büyük ve çeşitli veri setleri ile model performansının iyileştirilmesi üzerine odaklanılabilir.

Bu çalışma, kredi kartı onay süreçlerinde makine öğrenimi modellerinin etkili bir şekilde kullanılabilmesi için bir temel oluşturmuştur. İleriye dönük olarak, model geliştirmeleri ve yeni özellik eklemeleri ile daha güçlü ve güvenilir sınıflandırıcılar elde etme potansiyeli bulunmaktadır.

Kaynaklar

[1] University of California Irvine Repository, Credit Card Approval Dataset
<https://archive.ics.uci.edu/dataset/27/credit+approval>

Son Erişim Tarihi: 03.12.2023

[2] Arun, K., Ishan, G., & Sanmeet, K., (2016). Loan approval prediction based on machine learning approach. *IOSR J. Computer Eng*, 18 (3), 18-21.

[3] Serengil, S. I. & Imece, S. & Tosun, U. G. & Buyukbas, E. B., & Koroglu, B., (2021). A Comparative Study of Machine Learning Approaches for Non Performing Loan Prediction. *In 2021 6th International Conference on Computer Science and Engineering (UBMK) (pp. 326- 331). IEEE.*

[4] Ali Tezcan Sarızeybek. & Onur Seveli. (2022). Makine Öğrenmesi Yöntemleri ile Banka Müşterilerinin Kredi Alma Eğiliminin Karşılaştırmalı Analizi. *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 5(2), 137–144.
<https://dergipark.org.tr/en/download/article-file/2129957>

[5] Alan, A., (2020). Makine öğrenmesi sınıflandırma yöntemlerinde performans metrikleri ile test tekniklerinin farklı veri setleri üzerinde değerlendirilmesi (Master's thesis, Fen Bilimleri Enstitüsü)

Gök M., (2017). Makine Öğrenmesi Yöntemleri ile Akademik Başarının Tahmin Edilmesi. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 5 (3), 139-148.

Burges, C. J., (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), 121-167.

Biau, G., & Scornet, E., 2016. A random forest guided tour. *Test*, 25 (2), 197-227.

Ali, J., Khan, R., Ahmad, N., & Maqsood, I., (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, 9(5), 272.