Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi
Pamukkale University Journal of Engineering Sciences

Veri madenciliği yöntemleriyle kredi skor tahmini: performans karşılaştırması ve analizi

Predicting credit scores with data mining methods: performance comparison and analysis

Deniz Kızılaslan[1], Hakan Burak Emekli[2]*

[1]Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği, İstanbul Aydın Üniversitesi, İstanbul, Türkiye.
denizkizilaslan@stu.aydin.edu.tr
[2]Anadolu BİL Meslek Yüksekokulu, Bilgisayar Programcılığı, Üniversitesi/ İstanbul Aydın Üniversitesi, İstanbul, Türkiye.
hakanemekli@aydin.edu.tr

Geliş Tarihi/Received: 03.10.2024 Düzeltme Tarihi/Revision: 07.08.2025 doi: 10.5505/pajes.2025.84577
Kabul Tarihi/Accepted: 10.09.2025 Araştırma Makalesi/Research Article

Öz

Kredi skoru tahmini, finansal kuruluşların kredi riskini etkin şekilde yönetmeleri ve sürdürülebilir kârlılık sağlamaları açısından kritik bir öneme sahiptir. Sağlıklı kredi kararlarının alınabilmesi için geçmiş verilere dayalı tahmin modellerinin geliştirilmesi gerekmektedir. Bu çalışmada, kredi notu veri seti üzerinde çeşitli makine öğrenmesi algoritmaları ile birlikte birliktelik kuralı çıkarımına dayalı Apriori algoritması kullanılarak tahmin modelleri oluşturulmuştur. Modelleme sürecinde veri madenciliği ve yapay zekâ tekniklerinden yararlanılmış; farklı sınıflandırma algoritmalarının başarı performansları 10 katlı çapraz doğrulama yöntemi ile doğruluk, hassasiyet (precision), hatırlama (recall) ve F1-skoru gibi metrikler üzerinden değerlendirilmiştir. İstatistiksel analizler (Wilcoxon ve paired t-testi) DNN modelinin Logistic Regression, Naive Bayes ve Random Forest modellerine göre anlamlı şekilde üstün olduğunu ortaya koyarken, MLP, SVM ve XGBoost modelleri ile benzer performanslar sergilediğini göstermiştir. Bu bulgu, DNN'nin özellikle karmaşık veri setlerinde güçlü bir tahmin modeli olduğunu desteklemektedir. Elde edilen sonuçlar, veri tabanlı tahmin yaklaşımlarının kredi risk analizinde etkinliğini göstermekte ve senaryoya uygun algoritma seçiminin önemine vurgu yapmaktadır. Ayrıca, kullanılan algoritmaların avantajları ve sınırlılıkları değerlendirilmiş; uygulama bağlamına göre en uygun yöntemin seçilmesinin kritik olduğu sonucuna varılmıştır. Bu çıktılar, kredi risk tahminine yönelik modelleme çalışmalarına katkı sağlamakta ve finansal kuruluşların karar destek sistemleri için yapay zekâ temelli çözümler geliştirmelerine rehberlik etmektedir.

Anahtar kelimeler: Madenciliği, Sınıflandırma algoritmaları, Kredi skor tahmin, Birliktelik analizi, Hiperparametre optimizasyonu.

Abstract

Credit scoring prediction is critically important for financial institutions to effectively manage credit risk and ensure sustainable profitability. Accurate credit decisions require the development of predictive models based on historical data. In this study, predictive models were developed using various machine learning algorithms along with the Apriori algorithm based on association rule mining applied to a credit scoring dataset. The modeling process leveraged data mining and artificial intelligence techniques; the performance of different classification algorithms was evaluated using 10-fold cross-validation through metrics such as accuracy, precision, recall, and F1-score. Statistical analyses (Wilcoxon signed-rank test and paired t-test) revealed that the Deep Neural Network (DNN) model outperformed Logistic Regression, Naive Bayes, and Random Forest models significantly, while exhibiting similar performance to MLP, SVM, and XGBoost models. This finding supports the strength of DNN models, particularly on complex datasets. The results demonstrate the effectiveness of data-driven predictive approaches in credit risk analysis and emphasize the importance of selecting algorithms appropriate to the scenario. Furthermore, the advantages and limitations of the applied algorithms were evaluated, concluding that choosing the most suitable method according to the application context is critical. These findings contribute to credit risk prediction modeling efforts and provide guidance for financial institutions in developing AI-based decision support systems.

Keywords: Data mining, Classification algorithms, Credit score estimation, Association analysis, Hyperparameter tuning.

1 Giriş

Veri madenciliği, büyük veri setlerinin analizinden elde edilen sonuçları içeren bir disiplindir. Bu disiplin, verileri işleyerek anlamlı hale getirmek ve bilgiye dönüştürmek için çeşitli teknikler kullanır. Veri madenciliği sürecinde istatistik, makine öğrenimi, yapay zekâ ve veri tabanı yönetimi gibi farklı alanlardan gelen yöntemler ve teknikler kullanılır. Büyük veri setlerinin analizi ve işlenmesi konusunda veri madenciliği son derece önemli bir araçtır, bu nedenle işletmeler, araştırmacılar ve diğer kuruluşlar veri madenciliği tekniklerini kullanarak verilerini analiz eder ve anlamlı bilgiler elde eder.

Finansal dünyada büyük bir öneme sahip olan kredi notu, ödeme geçmişine dayalı olarak belirlenen bir puanlama sistemidir ve kredi verme işlemlerinde kullanılır. Kredi skoru tahmini, bir bireyin veya kuruluşun kredi skorunu belirlemek için kullanılan bir yöntemdir.

Veri madenciliği, kredi skoru tahmininde büyük bir öneme sahiptir. Veri madenciliği teknikleri, büyük veri setlerinin analiz edilmesiyle elde edilen verileri işleyerek, kredi skoru tahmini için kullanılacak modellerin oluşturulmasını sağlar. Kredi skoru tahmini için kullanılan veriler arasında ödeme geçmişi, borç oranı, gelir seviyesi gibi faktörler yer alır. Bu faktörlerin analizi, veri madenciliği yöntemleriyle

*Yazışılan yazar/Corresponding author

gerçekleştirilir ve bu sayede bir bireyin veya kuruluşun kredi notu tahmini yapılmış olur.

Bu makalede, kredi notu tahmini yapmak için veri madenciliği yöntemlerinin kullanılmasının önemi ve faydaları vurgulanacaktır. Ayrıca, kredi notu tahmini için kullanılan yöntemler ve bu yöntemlerin performanslarının karşılaştırılması da ele alınacaktır.

2 Veri madenciliği teknikleri

Veri madenciliği, büyük ve çeşitli veri ambarlarından daha önce keşfedilmemiş bilgileri ortaya çıkarmak ve bunları karar verme ve eylem planlaması için kullanmak için kullanılan bir süreçtir. Bu süreç, büyük veri kümeleri içinde gelecekle ilgili tahminler yapmak için ilişkileri ve kuralları aramayı içerir. Veri madenciliği, verilerin desenlerini, ilişkilerini, değişimlerini, düzensizliklerini, kurallarını ve istatistiksel olarak önemli yapılarını yarı otomatik olarak keşfetme işlemidir. Bilgisayar, veriler arasındaki ilişkileri, kuralları ve özellikleri belirlemekten sorumludur [1]. Temel hedef, daha önce fark edilmemiş veri desenlerini tespit etmektir. Etkili bir veri madenciliği uygulaması için farklı veri tipleriyle çalışabilme, veri madenciliği algoritmalarının etkinliği ve ölçeklenebilirliği, sonuçların yararlılık, kesinlik ve anlamlılık kriterlerini karşılaması, keşfedilen kuralların çeşitli şekillerde gösterilebilmesi, farklı ortamlardaki veriler üzerinde işlem yapabilmesi, gizlilik ve veri güvenliği gibi özelliklerin sağlanması gerekmektedir [2]. Veri madenciliği aslında bilgi keşfinin bir parçası olarak kabul edilir ve bilgi keşfi sürecindeki aşamaları içerir.

Veri Temizleme, veri içerisinde bulunan gürültülü ve tutarsız verilerin temizlenmesi

Veri Bütünleştirme, farklı kaynaklardan elde edilen verilerin birleştirilerek tek bir veri kaynağına entegre edilmesidir.

Veri Seçme, veri setindeki gereksiz veya önemsiz verilerin elenmesini ve sadece analiz veya modele katkı sağlayacak önemli verilerin belirlenmesidir.

Veri Dönüşümü, veri setinde yer alan verilerin işlenmesi ve dönüştürülmesini içerir. Bu adım, verileri daha uygun bir şekilde analiz etmek, modellemek veya kullanmak için değiştirme ve uyarlanmadır [3,4].

Veri madenciliği için hem ticari hem de açık kaynak programlar mevcuttur. Bu programlarda birçok farklı algoritma bulunur ve bu algoritmalar kullanılarak eldeki verilerden anlamlı bilgiler çıkarılabilir.

Bu başlık altında, veri madenciliği sürecinde kullanılan temel teknikleri inceleyeceğiz. Veri madenciliği, büyük veri kümelerinden değerli bilgileri çıkarmak amacıyla çeşitli yöntemleri içerir.

Desen tanıma (Pattern recognition)

Desen tanıma, verilerdeki düzenleri ve tekrarlanan yapıları tanımlama sürecidir. Bu düzenler, görsel, işitsel, metinsel veya diğer veri türlerinde bulunabilir. Desen tanıma, bir desenin veri içinde tekrarlandığını belirlemek ve bu desenlerin anlamını çıkarmak için kullanılır [5].

Sınıflandırma (Classification)

Sınıflandırma, verileri belirli bir kategoriye sınıflandırma işlemidir. Bu işlem, öğrenilmiş bir model kullanılarak verinin hangi sınıfa ait olduğunu belirler. Örnek uygulamalar arasında spam e-posta tespiti, tıbbi teşhisler ve görüntü tanıma bulunur [6].

2.1 Kümeleme (Clustering)

Kümeleme, verileri benzer özelliklere sahip gruplara bölmeyi amaçlar. Veriler arasındaki doğal yapıyı tanımak için kullanılır. Kümeleme, pazar segmentasyonu, müşteri profilleri ve coğrafi alanlar gibi birçok uygulamada kullanılır [7].

2.2 Tahmin ve regresyon (Prediction and Regression)

Tahmin ve regresyon, gelecekteki değerleri tahmin etmek veya birçok değişken arasındaki ilişkiyi modellemek için kullanılır. Örneğin, hava durumu tahmini, hisse senedi fiyatları tahmini ve enerji tüketimi regresyonu gibi uygulamalar bu kategoriye girer [8].

2.3 Derin öğrenme (Deep learning)

Derin öğrenme, yapay sinir ağları kullanarak çok katmanlı ve karmaşık veri temsilleri öğrenmeyi hedefler. Derin öğrenme, büyük veri kümelerindeki desenleri tanımlamak ve karmaşık görevleri otomatik olarak gerçekleştirmek için kullanılır. Örnek uygulamalar arasında görüntü sınıflandırma, metin analizi ve ses tanıma bulunur [9].

Bu yöntemler, veri madenciliği uygulamalarında yaygın bir şekilde kullanılır. Her bir yöntem, özel veri analiz ihtiyaçlarını karşılamak için farklı bir işlevi yerine getirir.

3 Literatür taraması

Makine öğrenmesi ve derin öğrenme yöntemlerinin kredi skoru tahmini alanında kullanımı son yıllarda önemli ölçüde artmıştır. Bu alandaki literatür, hem geleneksel makine öğrenimi algoritmaları hem de modern derin öğrenme yaklaşımlarını kapsamlı biçimde incelemekte ve farklı veri setleri üzerinde algoritmaların performanslarını karşılaştırmaktadır. Bu çalışmanın "Ön Çalışmalar" başlığı altında, mevcut literatür özetlenerek bu çalışmanın yeri ve katkısı ortaya konmaya çalışılmıştır.

Daha önceki sistematik incelemeler, kredi riskinin değerlendirilmesinde makine öğrenimi tekniklerinin kullanımına dair genel bir bakış sağlamış ve farklı algoritmaların kredi risk tahminindeki etkilerini detaylı şekilde analiz etmiştir [10,11]. Lojistik regresyon, destek vektör makineleri (SVM), karar ağaçları, rastgele ormanlar (Random Forest), sinir ağları gibi yaygın kullanılan algoritmaların yanı sıra, son dönemlerde XGBoost ve çeşitli derin öğrenme modelleri de popülerlik kazanmıştır. Bu algoritmaların kredi riski tahminindeki doğruluk, precision, recall, F1-score gibi performans metrikleri üzerinden karşılaştırmaları yapılmıştır.

Literatürde farklı makine öğrenimi tekniklerinin değişken başarı oranlarına sahip olduğu, belirli veri seti ve uygulama koşullarında bazı algoritmaların diğerlerine kıyasla üstün performans gösterdiği ancak genel anlamda tek bir “en iyi” algoritmanın olmadığı yönünde görüş birliği mevcuttur [10,11]. Bu durum, uygulama senaryosuna ve kullanılan veri setine göre algoritma seçiminin kritik önem taşıdığını göstermektedir. Ayrıca, eğitim süreleri, model karmaşıklığı ve uygulama pratikliği gibi faktörlerin de karar verme sürecinde göz önünde bulundurulması gerektiği vurgulanmaktadır.

Örnek olarak, Brown ve ark. [12] karar ağaçları ile destek vektör makinelerini kıyaslamış ve SVM’nin küçük veri setlerinde daha iyi sonuç verdiğini raporlamıştır. Chen ve Zhang [13], XGBoost algoritmasını kullanarak büyük ölçekli kredi verisi üzerinde %92’lik F1-score elde etmişlerdir. Li ve ark. [14], sinir ağlarının yüksek doğruluk sağlamakla beraber eğitim sürelerinin uzun olması nedeniyle pratik uygulamalarda zorluklar yaratabileceğini belirtmiştir. Zhou ve arkadaşları

[15], derin öğrenme tabanlı modellerin (DNN ve LSTM) klasik makine öğrenimi yöntemlerine göre daha iyi genelleme kabiliyetine sahip olduğunu ileri sürerken, Ghosh ve ark. [16] lojistik regresyon gibi basit algoritmaların belirli koşullarda daha stabil sonuçlar verdiğini ortaya koymuştur. Benzer şekilde, Zontul, Polat ve Yıldırım (2021) tarafından yapılan bir çalışmada da farklı makine öğrenmesi algoritmalarının kredi skoru tahmininindeki performansları karşılaştırılmış ve özellikle Random Forest ile XGBoost algoritmalarının yüksek başarı sağladığı belirtilmiştir [17].

Buna ek olarak, son yıllarda yapılan çalışmalar da benzer veri setleri ve algoritmalar üzerinden performans karşılaştırmalarına odaklanmıştır. Örneğin, Arram ve ark. (2023) Vietnam'daki öğrenci kredi skoru tahmininde Random Forest, Gradient Boosting, SVM ve DNN modellerini karşılaştırmış ve her modelin farklı güçlü yönlere sahip olduğunu raporlamıştır [18]. Hussain ve ark. (2023) ise geleneksel kredi skoru teknikleri ile AI tabanlı yaklaşımları karşılaştırarak, yapay zeka yöntemlerinin model doğruluğunu artırabileceğini göstermiştir [19]. Mhlanga (2023) çalışmasında ise İslami ve geleneksel bankalarda kredi risk tahmini için CatBoost ve diğer makine öğrenimi algoritmalarının performansını incelemiş ve CatBoost'un yüksek doğruluk sağladığını bildirmiştir [20].

Bu çalışmaların büyük çoğunluğunda algoritmaların doğruluk, precision, recall ve F1-score gibi temel performans metrikleriyle değerlendirilmesine rağmen, eğitim süresi ve hesaplama maliyeti gibi pratik kriterlerin çoğunlukla göz ardı edildiği görülmektedir. Ayrıca, birçok çalışma sadece belirli birkaç algoritmayı veya sınırlı sayıda veri setini kullanmakla sınırlı kalmıştır.

Bu çalışma ise, literatürdeki bu boşluğu doldurmak amacıyla, kredi skoru tahmini için geniş bir yelpazede makine öğrenimi algoritmalarını (Logistic Regression, Naive Bayes, J48 (Decision Tree), Random Forest, SVM, MLP, XGBoost, SNN, DNN) ve apriori algoritmasıyla birliktelik kuralı çıkarımını karşılaştırmalı olarak değerlendirmiştir. Veri setindeki kategorik değişkenler LabelEncoder ile dönüştürülmüş, model eğitimlerinde scikit-learn, TensorFlow ve Keras kütüphaneleri kullanılmıştır. Ayrıca, modellerin eğitim süreleri ve performansları birlikte analiz edilerek, doğruluk kadar uygulama pratikliği de göz önünde bulundurulmuştur.

Bu kapsamlı değerlendirme, kredi riski tahmininde hem doğruluk hem de işlem süresi açısından algoritmaların avantaj ve dezavantajlarını ortaya koyarak, finansal kurumların karar destek sistemleri için daha bilinçli algoritma seçimi yapmasına yardımcı olmayı amaçlamaktadır. Böylece çalışma, literatürde genellikle göz ardı edilen eğitim süresi ve algoritma çeşitliliği boyutlarında önemli katkılar sunmaktadır.

Sonuç olarak, hem sistematik inceleme hem uygulamalı analiz yönleriyle, kredi risk tahmininde makine öğrenimi tekniklerinin etkinliği ve algoritma seçiminin önemi bu çalışma ile detaylı biçimde ortaya konmuştur.

4 Veri ön işleme aşamaları

Çalışmada, 2020-2022 yılları arasında Türkiyede bulunan bir finans kuruluşuna ait ve toplam 88.624 veri noktasından oluşan bir veri kümesi kullanılmıştır. Veri ön işleme süreçleri sonucunda, anlamlı ve eksiksiz olan 57.850 veri ile analizlere devam edilmiştir. Kişilerin cinsiyeti, eğitim durumu, yaşı, medeni hali, gelir düzeyi ve devam eden kredi durumuna göre kredi skoru tahminlemesi gerçekleştirilmiştir. Elde edilen sonuçların, kredi başvuru değerlendirme süreçlerinde ilgili finans kuruluşuna yol gösterici olması hedeflenmektedir.

4.1 Veri seti

Bu çalışmada, kredi skoru üzerinde etkili olduğu düşünülen değişkenler belirlenmiştir. Belirlenen değişkenler kredi alıcılarına ait çeşitli nitelikleri içermektedir. Belirlemiş olduğumuz değişkenlere ait bilgilere aşağıda değinilmektedir.

Cinsiyet: Araştırmalar, kadınların borç ödeme konusunda erkeklere oranla daha düzenli olduklarını göstermektedir. Bu araştırmalar dikkate alınarak, cinsiyetin kredibilite üzerinde etkili bir faktör olarak değerlendirilmiş ve çalışmamızda kullanılmıştır [21].

Eğitim Durumu: Araştırmalar, eğitim düzeyinin kredibilite üzerinde önemli bir etkisi olduğunu göstermektedir. Daha yüksek eğitim seviyesine sahip bireylerin, finansal sorumluluklarını daha iyi yönetme eğiliminde oldukları ve bu nedenle kredi ödemelerinde daha düzenli oldukları gözlemlenmiştir [22]. Bu nedenle, çalışmamızda eğitim durumu, kredibiliteyi belirleyen önemli bir faktör olarak ele alınmıştır.

Yaş: Yaş kredibiliteyi etkileyen bir faktör olarak çalışmada kullanılmıştır.

Gelir Düzeyi: Gelir düzeyi, bireylerin kredi ödemelerini düzenli bir şekilde yapabilme kapasitesini doğrudan etkileyen önemli bir faktördür. Yüksek gelirli bireylerin, finansal yükümlülüklerini daha sorunsuz yerine getirebilme ihtimali göz önünde bulundurularak çalışmamızda gelir düzeyi, kredibilitenin değerlendirilmesinde kritik bir değişken olarak kullanılmıştır.

Takibi Devam Eden Kredi Durumu: Takibi devam eden kredi durumu, bir bireyin mevcut kredi borçlarını ne kadar düzenli ödediğini ve bu borçların ne kadarını hala ödemekte olduğunu gösteren önemli bir göstergedir. Bu durum, bireyin finansal disiplinini ve sorumluluklarını yerine getirme kapasitesini yansıtır. Düzenli ödeme yapmayan veya borçlarını aksatan bireyler, genellikle daha yüksek kredi riski taşırlar. Bu nedenle, çalışmamızda takibi devam eden kredi durumu, kredibiliteyi belirleyen kritik faktörlerden biri olarak ele alınmıştır [23].

4.2 Veri temizleme

Veri temizleme, veri madenciliği sürecinin en önemli ve zaman alıcı adımıdır. Veri kümesindeki Gelir Düzeyi, Takibi Devam Eden Kredi Sayısı, Eğitim Durumu, Yaş gibi kategorik veriler yüksek miktarda eksik değer içermektedir. Kategorik verilerdeki eksik değerlerin çözümü için, eksik değerlerin yerine atama yöntemleri kullanılabilir. Eğer bir özniteliğe ait verilerin büyük bir çoğunluğu eksikse, o öznitelik veri kümesinden çıkarılmalıdır.

Kategorik verilerdeki eksik değerlerin çözümü için,mod atama yöntemleri kullanılmıştır. Eğer bir özniteliğin büyük bir kısmında eksik veri mevcutsa, bu öznitelik veri setinden tamamen çıkarılmıştır. Sayısal (numerik) öznitelikler için ise eksik değerlerin yerine ortalama (mean) atama yöntemi uygulanmıştır. Bu süreçte Python'da pandas kütüphanesi kullanılmış ve veri setindeki eksik değerler titizlikle yönetilmiştir. Veri temizleme işlemi sonucunda, veri setimizde 88.624 veriden 57.850 adet anlamlı veriye ulaşılmıştır. Niteliksiz veriler, veri setinden çıkarılarak, elde edilecek sonuçların doğruluğu artırılmıştır.

4.3 Veri dönüştürme

Veri dönüştürme işlemi, makine öğrenimi projelerinde model performansını artırmak için önemlidir. Python kütüphaneleri kullanılarak gerçekleştirilen veri dönüştürme adımları, verilerin daha uygun bir formata getirilmesini sağlamaktadır. Bu süreçte, özellikle pandas ve sklearn gibi kütüphaneler sıklıkla kullanılmaktadır.

İlk olarak, kategorik değişkenlerin nominal veriye dönüştürülmesi için label encoding ve one-hot encoding yöntemleri uygulanmıştır.Makine öğrenimi modellerinin kategorik değişkenler ile daha iyi çalışabilmesi amacıyla Takibi Devam Eden Kredi Sayısı, Eğitim Durumu, Gelir Düzeyi, Yaş gibi kategorik değişkenler, Python'da LabelEncoder ve OneHotEncoder kullanılarak nominal ve sayısal verilere dönüştürülmüştür. Son olarak, OneHotEncoder yardımıyla kategorik değişkenlerin her bir kategorisi için ayrı bir sütun oluşturulmuştur.

Veri dönüştürme sürecinde sayısal (numerik) veriler de normalizasyon işlemine tabi tutulmuştur. Normalizasyon, sayısal verilerin belirli bir ölçeğe getirilmesini sağlar ve modelin farklı ölçeklerdeki verilerle daha etkin bir şekilde çalışmasına olanak tanır. Python'da sayısal verilerin normalizasyonu için sklearn kütüphanesinin MinMaxScaler ve StandardScaler sınıfları kullanılmıştır. Veri dönüştürme süreci tamamlandıktan sonra, veri seti modellemeye hazır hale gelmiş ve modelin performansının arttırılmasına katkı sağlamıştır.

4.4 Korelasyon matrisi

Korelasyon matrisi, bir veri setindeki değişkenler arasındaki doğrusal ilişkileri ve bu ilişkilerin gücünü gösteren simetrik bir tabloyu ifade etmektedir. Matrisin her hücresi, iki değişken arasındaki korelasyon katsayısını temsil etmektedir. Bu katsayılar -1 ile +1 arasında değişmekte ve ilişkinin yönünü ve gücünü belirtmektedir [24].

+1: Mükemmel pozitif korelasyon.İki değişken arasında mükemmel pozitif doğrusal ilişki olduğunu ifade etmektedir. Bir değişken artarken diğer değişken de aynı oranda artmaktadır.

-1: İki değişken arasında mükemmel negatif doğrusal ilişki olduğunu ifade etmektedir.Bir değişken artarken diğer değişken azalmaktadır.

0: İki değişken arasında doğrusal bir ilişki yoktur anlamını taşımaktadır.

Veri setimiz üzerinde elde etmiş olduğumuz kolerasyon matrisine ait bilgiler Şekil 1' de sunulmuştur.

Elde edilen korelasyon matrisine ait analiz sonuçları aşağıda detaylı olarak ele alınmıştır.

Cinsiyet ile diğer değişkenler arasında belirgin bir korelasyon gözlemlenmemektedir. Cinsiyet ve Devam Eden Kredi arasında pozitif ama çok zayıf bir korelasyon vardır (0.06)

Eğitim Durumu ile Yaş Skalası arasında negatif bir korelasyon vardır (-0.283). Bu, eğitim düzeyi arttırça yaşın genellikle azaldığını ifade etmektedir. Diğer değişkenler çok düşük korelasyonlara sahiptir.

Gelir Düzeyi ile Yaş Skala arasında pozitif bir korelasyon var (0.105). Bu, yaş arttıkça gelir düzeyinin de genellikle arttığını göstermektedir.

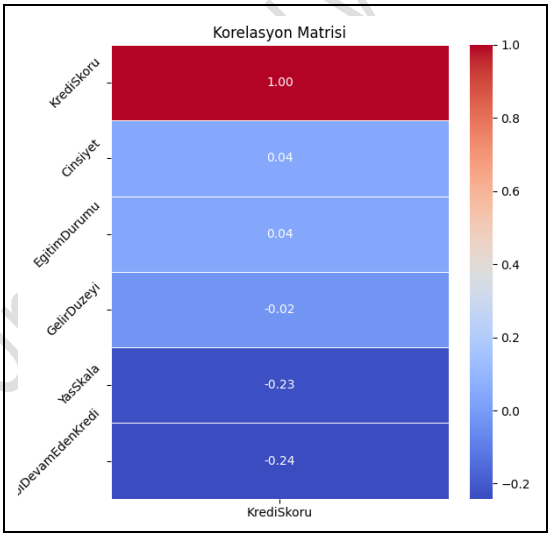
Gelir Düzeyi ve Takibi Devam Eden Kredi arasındaki pozitif korelasyon (0.121), gelir düzeyi arttıkça, bireylerin takibi devam eden kredi sayısının da hafif bir artış gösterebileceğini göstermektedir.

Yaş Skala ile Eğitim Durumu arasında negatif bir korelasyon var (-0.283).Bu, yaş skalası arttıkça eğitim düzeyinin düşmekte olduğunu ifade etmektedir.

Yaş Skala ve Kredi Skoru arasında negatif bir korelasyon var (-0.228). Bu, yaş arttıkça kredi skorunun azaldığını ifade etmektedir.

Yaş Skala ve Takibi Devam Eden Kredi arasındaki pozitif korelasyon (0.046), bu iki değişken arasında çok zayıf bir doğrusal ilişki olduğunu göstermektedir. Pozitif bir korelasyon olması, Yaş Skalası arttıkça Takibi Devam Eden Kredi sayısının da artma eğiliminde olduğunu belirtmektedir.

Kredi Skoru, diğer değişkenlerle genellikle düşük korelasyonlar göstermektedir.



Şekil 1. Korelasyon matrisi.

Figure 1. Correlation matrix.

4.5 Test kümesi seçimi

Algoritmaları çalıştırma esnasında test yöntemi olarak kullanılan method "10-kat çapraz doğrulama" metodudur. Bu method sayesinde veri kaynağımız 10 bölüme ayrılmaktadır. Ayrılan her bölüm bir defa test kümesi olarak, diğer 9 bölüm ise öğrenme kümesi olarak kullanılmaktadır.

Veri kümesi üzerinde farklı modeller deneyerek en yüksek doğruluk oranına sahip olan model seçilmiştir.

Tablo 1' de veri setimiz içerisinde bulunan nitelikler, niteliklere ait özellikler ve özelliklerin veri seti içerisinde bulunan kayıt sayıları listelenmektedir.

Tahminleme ve başarım ölçütlerini karşılaştırmak amacıyla Logistic Regression,Naive Bayes, J48, RandomForest, SVM(Support Vector Machine), MLP (Multi-Layer Perceptron), XGBoost, SNN (Simple Neural Network ve DNN (Deep Neural Network) algoritmalarından faydalanılmıştır. Veri seti, eğitim ve test verisi olmak üzere ikiye bölünerek modelin performansı değerlendirilmiştir. Eğitim seti, modelin öğrenme sürecinde kullanıldı ve test seti, modelin tahmin gücünü değerlendirmek için ayrılmıştır. Bu ayrım, modelin henüz görmediği veriler üzerinde performansını test etmemizi sağlamaktadır. Modelin %80'lik kısmı eğitim verisi, %20'lik kısmı ise test verisi olarak kullanılmıştır.

Tablo 1. Veri seti istatistikleri.

Table 1. Dataset Statistics.

Nitelik	Özellik	Veri sayısı
Cinsiyet	Kadın	17.433
	Erkek	40.417
Eğitim durumu	İlkokul	13.232
	Ortaokul	10.938
	Lise	21.278
	Lisans	1.375
	Yüksek lisans ve üzeri	11.927
Gelir düzeyi	Normal	34.249
	Yüksek	23.601
Yaş aralığı	Genç	20.740
	Orta yaş	26.611
	Yaşlı	10.499
Takibi devam eden kredi durumu	Var	8.306
	Yok	49.544

5 Denetimli sınıflandırma algoritmaları ve yapay sinir ağları

Denetimli sınıflandırma algoritmaları, verileri önceden belirlenmiş sınıflara atama görevini yerine getiren önemli araçlardır. Bu makalede, bazı yaygın denetimli sınıflandırma algoritmalarını ve güncel yapay sinir ağı modellerini ele alacağız: Logistic Regression,Naive Bayes,J48, RandomForest,SVM(Support Vector Machine),MLP (Multi-Layer Perceptron),XGBoost,SNN (Simple Neural Network ve DNN (Deep Neural Network) algoritmaları üzerinde tahminler yaparak doğruluk karşılaştırması yapılması amaçlanmaktadır.

5.1 Naive bayes

Naive Bayes, olasılık temelli bir sınıflandırma algoritmasıdır ve Bayes Teoremi'ne dayanır. Temel varsayım, veri özelliklerinin birbirinden bağımsız olduğudur (bu "naive" olarak adlandırılır). Bayes Teoremi, veri noktalarının sınıf olasılıklarını tahmin etmek için kullanılır. Özellikle metin sınıflandırma gibi uygulamalarda başarılıdır [5,25].

5.2 J48 algoritması

J48 veya C4.5, ağaç yapısı kullanarak sınıflandırma yapabilen bir karar ağacı algoritmasıdır. Veri kümesini adım adım böler ve enformasyon kazancını kullanarak en iyi bölünmeleri seçer. Sonuçta bir karar ağacı oluşturur. C4.5, hem sayısal hem de kategorik verilerle çalışabilir ve sınıflandırma ve regresyon problemleri için kullanılabilir [26].

5.3 Random forest

Rastgele Orman (RF), yaygın olarak kullanılan güçlü bir denetimli makine öğrenmesi algoritmasıdır. RF, içerdiği her karar ağacının bağımsız olarak rastgele örneklemlenen vektör değerlerine dayandığı ve bu ağaç tahmincilerinin kombinasyonuyla bir orman oluşturduğu bir ensemble (topluluk) yöntemidir. Diğer denetimli makine öğrenimi teknikleri gibi, RF eğitim ve değerlendirme aşamalarından oluşur. Temel olarak, bu algoritma, eğitim verilerinde bulunan etiketlerden yeni, etiketlenmemiş giriş verilerinin etiketlerini tahmin etmek için öngörülerde bulunur [27].

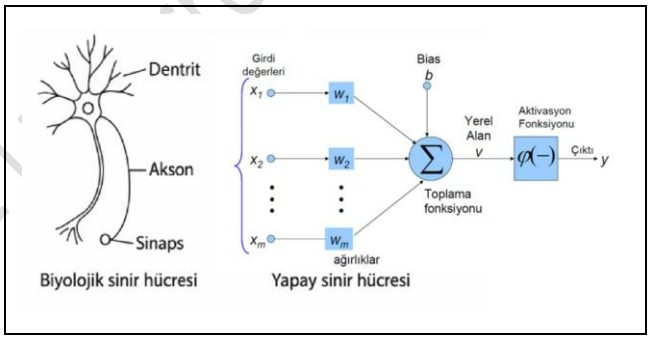
5.4 Destek vektör makinesi

Support Vector Machine (SVM), makine öğrenimi alanında kullanılan güçlü bir sınıflandırma ve regresyon algoritmasıdır. SVM, veri noktalarını sınıflandırmak için optimal bir karar sınırı oluşturur; bu sınır, veri noktalarını en iyi şekilde ayıran ve genelleme yeteneği en yüksek olan sınırdır. Temelde, sınıflar arasında bir marjin oluşturarak bu marjini maksimize eden bir karar sınırı bulmaya çalışır. Bu marjin içinde kalan ve sınıflar arasındaki ayrımı en iyi sağlayan veri noktaları destek vektörleri olarak adlandırılır. SVM, lineer ve non-lineer sınıflandırma problemleri için kernel yöntemlerini kullanarak geniş bir uygulama alanına sahiptir [28].

5.5 Yapay sinir ağları

MLP (Çok Katmanlı Perceptron), çok katmanlı yapay sinir ağlarına dayanan bir derin öğrenme algoritmasıdır. İnsan beyninin işleyişini taklit eden bir yapısı vardır ve çok karmaşık işlemleri gerçekleştirebilir. Derin öğrenme alanında büyük başarı elde etmiştir ve görüntü tanıma, doğal dil işleme gibi pek çok uygulama alanında kullanılır [29].

Warren McCulloch ve Walter Pitts, 1943 yılında "Sinir Aktivitesinde Düşüncelere Ait Bir Mantıksal Hesap" adlı makaleleriyle ilk yapay sinir ağı modelini tanıttılar.Yapay sinir ağları (YSA), biyolojik sinir ağlarını taklit eden sentetik yapılar olarak bilinir [30,31]. Biyolojik sinir hücresi yapısı ve yapay sinir ağı arasındaki benzerlik Şekil 2'de gösterilmektedir.



Şekil 2. Biyolojik sinir hücresi ve yapay sinir ağı [30].

Figure 2. Biological neuron and artificial neural network [30].

5.6 Lojistik regresyon

Logistic Regression (Lojistik Regresyon), makine öğrenmesinde sınıflandırma problemleri için kullanılan bir lineer modeldir. Sınıflandırma problemlerinde yaygın olarak kullanılmaktadır. Özellikle binary (iki sınıflı) sınıflandırma problemlerinde yaygın olarak tercih edilmektedir [31].

5.7 Aşırı gradyan arttırma (XGBoost)

XGBoost, yüksek performanslı ve esnek bir gradient boosting (gradyan artırma) kütüphanesidir. 2016 yılında Tianqi Chen tarafından geliştirilen bir makine öğrenme algoritmasıdır. XGBoost, gradyan artırma algoritmasını daha hızlı ve daha verimli bir şekilde uygulamak için çeşitli optimizasyon teknikleri kullanmaktadır. Paralel işleme ve donanım optimizasyonları sayesinde diğer gradyan artırma algoritmalarına göre daha hızlıdır. Verimli bir şekilde overfitting'i (aşırı öğrenme) azaltan düzenleme teknikleri kullanarak yüksek doğruluk sağlamaktadır [32].

5.8 Derin sinir ağları (DNN)

DNN (Derin Sinir Ağları), makine öğreniminde kullanılan bir tür yapay sinir ağıdır. Bu ağlar, birçok katmandan oluşur ve her bir katman, verilerdeki karmaşık desenleri öğrenmek için kullanılır. Derin sinir ağları, özellikle büyük veri setleri ve karmaşık problemler için güçlü bir modelleme aracıdır.

DNN'ler, giriş katmanı, bir veya daha fazla gizli katman ve çıkış katmanı olmak üzere birkaç katmandan oluşur. Gizli katmanlar, verilerdeki özellikleri öğrenerek modelin karmaşıklığını artırır. Bu ağlar, geriye yayılım (backpropagation) algoritması ve gradyan inişi (gradient descent) yöntemi ile eğitilir. Aşırı öğrenmeyi (overfitting) önlemek için Dropout, L1/L2 düzenleme gibi teknikler kullanılır.

DNN'ler, farklı optimizasyon algoritmaları kullanarak daha hızlı ve daha verimli öğrenme sağlar. Adam, RMSprop ve SGD (Stochastic Gradient Descent) yaygın olarak kullanılan optimizasyon algoritmalarıdır. Karmaşık veri desenlerini öğrenme ve yüksek doğrulukta tahminler yapma yetenekleri nedeniyle, DNN'ler modern makine öğreniminde yaygın olarak kullanılmaktadır [33].

5.9 Ateşleyen Sinir Ağları(SNN)

SNN (Spiking Neural Networks), biyolojik sinir ağlarının çalışma prensiplerinden ilham alarak tasarlanan yapay sinir ağı modelleridir. Bu tür sinir ağları, beyindeki sinir hücrelerinin (nöronların) gerçek zamanlı olarak ateşleme şeklini modellemeye çalışır. Genellikle giriş katmanı, gizli katman(lar) ve çıkış katmanı olmak üzere klasik yapay sinir ağı mimarisini takip eder. Ancak her katmandaki nöronlar ve aralarındaki bağlantılar, spiker aktivitesini ve zaman bağlamını hesaba katarak tasarlanır.

SNN'ler, genellikle biyolojik öğrenme prensiplerine dayanan eğitim algoritmaları ile eğitilir. Bu eğitim süreci, nöronların ateşleme zamanlarını ve aralarındaki sinaptik güçleri (bağlantı güçleri) optimize ederek gerçekleştirilir. Bu şekilde, sinir ağının daha doğru ve biyolojik olarak gerçekçi tepkiler vermesi amaçlanır [34].

6 Performans karşılaştırması ve metrikleri

Veri madenciliği bağlamında, model başarım ölçütleri oldukça kritiktir. Bu ölçütler, bir veri madenciliği modelinin performansını değerlendirmek için kullanılan metriklerdir.

6.1 Performans metrikleri

Performans metrikleri, veri madenciliği modellerinin performansını anlamak için kritik öneme sahiptir ve doğru metriklerin seçilmesi, modelin gerçek dünya problemlerine uygun şekilde değerlendirilmesini sağlamaktadır [6].

6.1.1 Doğruluk(Accuracy)

Doğruluk, bir modelin genel performansını değerlendirmede önemli bir ölçüttür. Fiziksel özellik ölçümünde, gerçek değer ile ölçülen değer arasındaki uyumsuzluk belirtilmiştir. Yüksek doğruluk oranı, modelin doğru sınıflandırma yaptığını gösterirken, düşük doğruluk oranı modelin yanlış sınıflandırma yaptığını belirtir. Bu metrik, modelin genel etkinliğini değerlendirirken kullanılır ve modelin performans güvenilirliği hakkında bilgi sağlar [35].

Doğruluk oranı yüksek olması modelin doğru bir şekilde sınıflandırıldığını, düşük olması ise modelin hatalı bir şekilde sınıflandırıldığını değerlendirmemizi sağlamaktadır. Doğruluk Oranı formülü Denklem (1) de tanımlanmıştır.

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

6.1.2 Başarı (Precision)

Precision (Başarı), beklenen sonucun doğruluğunu kesin olarak gösteren bir metriktir. Bu, bir modelin iddia ettiği gerçek pozitiflerin doğruluğunun, tüm pozitif tahminlerle ilişkisinin bir göstergesidir. Modelin talepleri konusunda hassas bir değerlendirme yapmak için kullanılır ve denklem (2) ile ifade edilir. Precision (Başarı) formülü Denklem (2) de tanımlanmıştır.

$$Başarı = \frac{TP}{TP + FP} \quad (2)$$

6.1.3 Hatırlama oranı(Recall)

Hatırlama Oranı(Recall), toplam pozitiflere odaklanır. Sistem, verilerdeki belirli sayıdaki kesin pozitifleri ifade etmektedir. Hatırlama Oranı(Recall) formülü Denklem (3) de tanımlanmıştır.

$$Hatırlama\ Oranı = \frac{TP}{TP + FN} \quad (3)$$

6.1.4 F1 skoru

F1 skoru, model performansını değerlendirmek için kullanılabilir. Hatırlama ve kesinlik değerlerinin harmonik ortalamasıdır F1 Skoru formülü Denklem (4)'te tanımlanmıştır.

$$F1\ Skor = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (4)$$

6.1.5 Karışıklık matrisi(Confusion matrix)

Sınıflandırma sonucu elde ettiğimiz tahmin sonuçlarının özetini sunan tablodur. Karışıklık matrisinde bulunan satırlar test kümesi içerisinde yer alan örneklere ait gerçek sayıları ifade ederken, matriste bulunan kolonlar modelin tahminlemesini ifade etmektedir.

		Var olan Durum	
		Pozitif Durumlar	Negatif Durumlar
Tahmin	Pozitif	TP	FP
	Negatif	FN	TN

Şekil 3. Karışıklık matrisi.

Figure 3. Confusion matrix.

Şekil 3'de sunulan Karışıklık matrisinden çıkan TP, FP, FN ve TN değerleri farklı metriklerle farklı performans ölçütlerine dönüştürülebilir ve hedeflediğimiz ölçütleri belirlememiz için bize yardımcı olmaktadır. TP, FP, FN ve TN değerleri aşağıda açıklanmıştır [32].

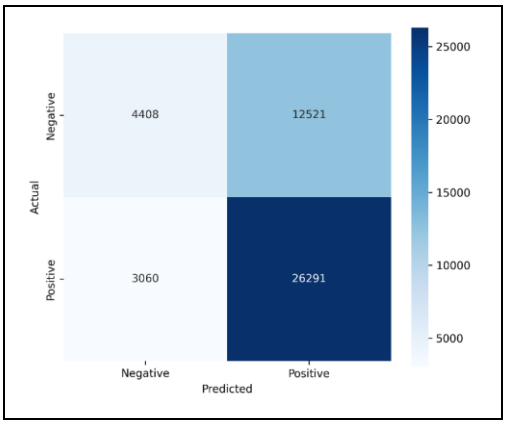
TP (Gerçek Pozitifler): Gerçek değer ve tahmin edilen değer Negatif olduğunda

(TN) Gerçek Negatifler: Gerçek değer ve tahmin edilen değer Negatif olduğunda

(FP) Yanlış Pozitifler: Gerçek değer negatif ancak tahmin değeri Pozitif olduğunda

(FN) Yanlış Negatifler: Gerçek değer Pozitif ancak tahmin değeri Negatif olduğunda.

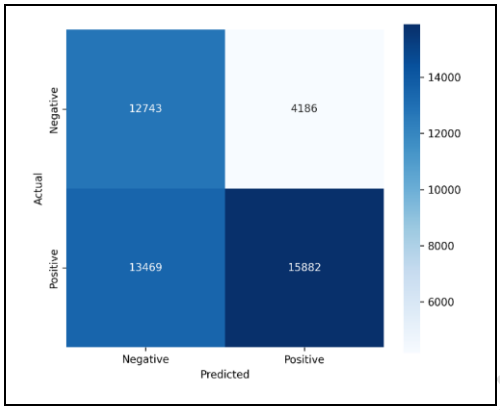
Şekil 4'te Logistic Regresyon modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 4. Logistic Regresyon Karışıklık matrisi.

Figure 4. Logistic Regression Confusion Matrix.

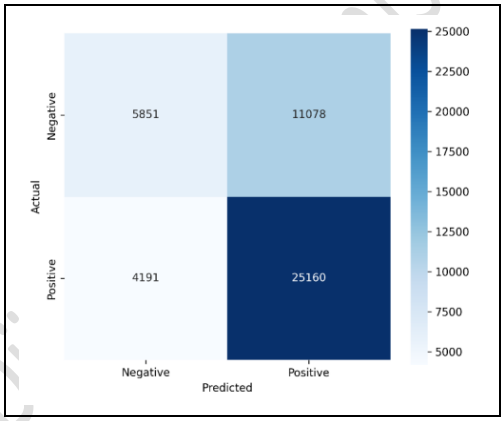
Şekil 5'te Naive Bayes modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 5. Naive Bayes Karışıklık matrisi.

Figure 5. Naive Bayes Confusion Matrix.

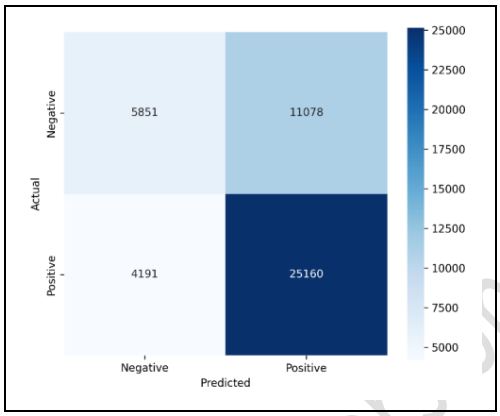
Şekil 6'da J48 modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 6. J48Karışıklık matrisi.

Figure 6. J48 Confusion Matrix.

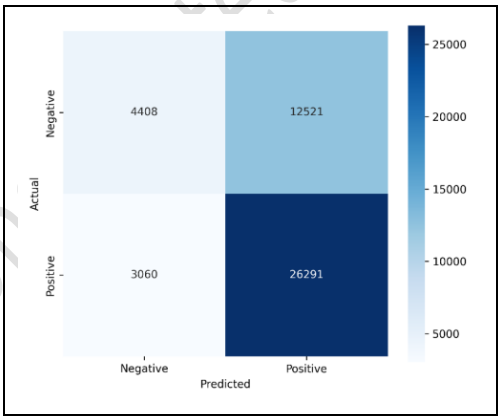
Şekil 7'de Random Forest modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 7.Random Forest Karışıklık matrisi.

Figure 7. Random Forest Confusion Matrix.

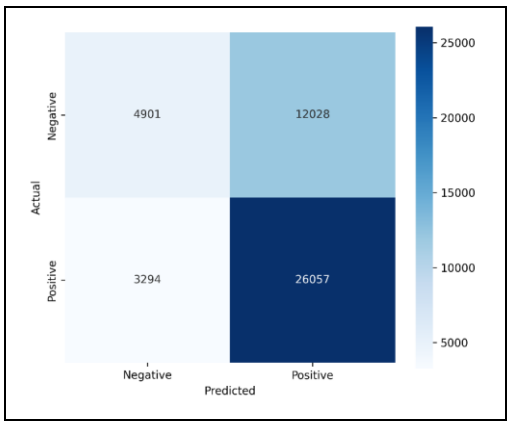
Şekil 8'de SVM modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 8. SVM Karışıklık matrisi.

Figure 8. SVM Confusion Matrix.

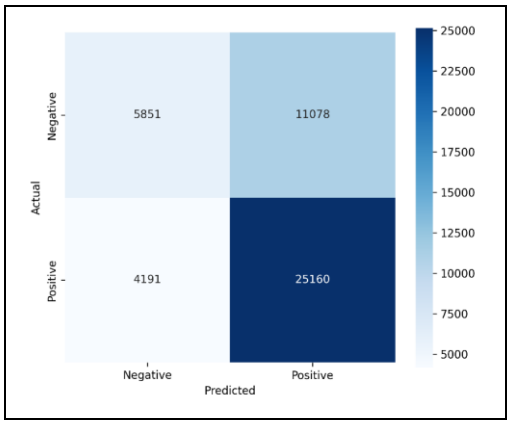
Şekil 9'da MLP modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 9.MLP Karışıklık matrisi.

Figure 9. MLP Confusion Matrix.

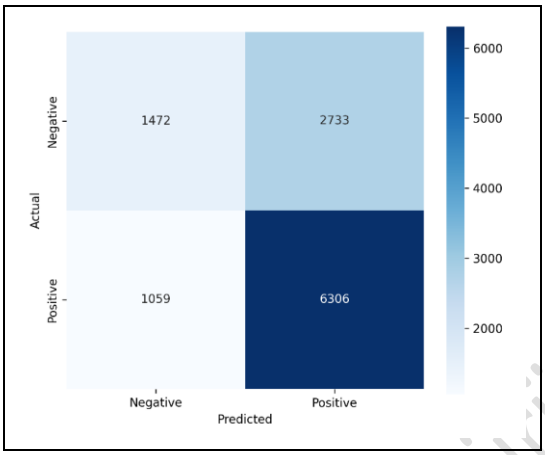
Şekil 10'da XGBoost modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 10. XGBoost Karışıklık matrisi.

Figure 10. XGBoost Confusion Matrix.

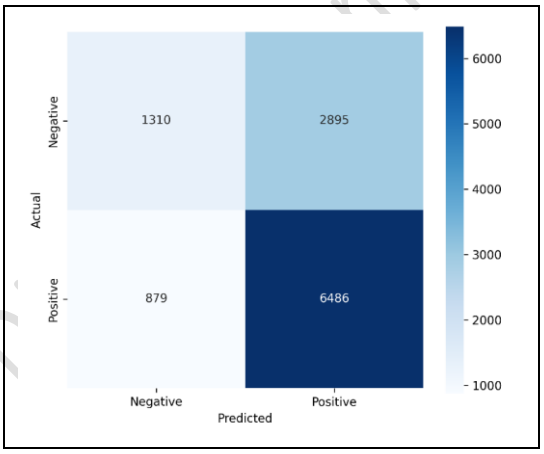
Şekil 11'de DNN modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 11.DNN Karışıklık matrisi.

Figure 11. DNN Confusion Matrix.

Şekil 12'de SNN modeline ait karışıklık matrisinden elde edilen değerler sunulmuştur.



Şekil 12.SNN Karışıklık matrisi.

Figure 12. SNN Confusion Matrix.

Aşağıda, her model için sağlanan karışıklık matrislerini yorumlayarak genel performans değerlendirmesi yapılmıştır. DNN, doğru pozitiflerin yüksekliği ve yanlış pozitiflerin nispeten düşük olması ile iyi performans gösteriyor. Ayrıca, F1-skora göre de en yüksek değerlere sahiptir. SNN, düşük doğru pozitif ve doğru negatif oranları ile diğer modellere göre daha düşük performans göstermektedir.

Farklı modellerin karışıklık matrislerini incelediğimizde, her bir modelin performansının çeşitliliği dikkat çekmektedir. Logistic Regression ve SVM modelleri benzer bir hata profili sergileyerek, pozitif sınıflarda yüksek yanlış pozitif oranları ve negatif sınıflarda yeterli doğruluk gösteriyor. Naive Bayes modeli, büyük oranda negatif sınıfları doğru sınıflandırırken, pozitif sınıflarda zayıf kalmakta, bu da modelin pozitif tahminlerde zayıf olduğunu göstermektedir. J48 ve Random Forest modelleri, her iki sınıfı da dengeli bir şekilde tahmin ederek, genel olarak güçlü bir performans sergilemektedir. MLP modeli, pozitif sınıflarda biraz daha iyi sonuçlar verirken, bazı yanlış negatif tahminler yapmaktadır. XGBoost ise, yüksek performansıyla her iki sınıfı da etkin bir şekilde ayırmakta ve genellikle doğru tahminler sağlamaktadır. DNN ve SNN modelleri, özellikle DNN, çok iyi bir performans sergilemekte ve pozitif sınıflarda daha yüksek başarı göstermektedir, ancak SNN modelinin performansı daha düşük kalmaktadır.

6.2 Performans Değerlendirmesi

En Yüksek Performans: DNN (Deep Neural Networks), en yüksek doğruluk (accuracy: 0.720709), hatırlama oranı (recall: 0.720709) ve F1 skoru (F1-Score: 0.680230) değerlerine sahip modeldir. Bu, DNN'nin genel performansının diğer algoritmalardan daha iyi olduğunu göstermektedir.

En İyi Kesinlik: Naive Bayes, kesinlik (precision: 0.680176) açısından en yüksek değeri sunmaktadır. Bu, modelin pozitif sınıfları doğru şekilde tanımlama yeteneğinin yüksek olduğunu gösterir.

Ortalama Performans:SVM (Support Vector Machine) ve XGBoost, yüksek doğruluk (accuracy: 0.716742 ve 0.719651) ve hatırlama oranı (recall: 0.716742 ve 0.719651) değerleri sunarak iyi performans gösterir, ancak F1 skorları DNN'ye göre daha düşüktür.

Kısa Sürede Çalışan Modeller: J48, Random Forest, SVM gibi modeller, accuracy ve F1 skoru bakımından orta seviyede performans gösterirken, daha hızlı eğitim sürelerine sahip olabilirler.

Üzerinde çalışılan modellerin performans karşılaştırılması yapılmış ve başarı oranı en yüksek algoritma DNN algoritması olarak belirlenmiştir.

Doğrulama ve tahmin için kullanılan verilerin metrikler ile oluşturulan modellere ait doğruluk, başarı, hatırlama oranı ve F1 skoru verileri Tablo 2'de sunulmuştur.

Tablo 3'te, Logistic Regression (LR), Naive Bayes (NB),J48, Random Forest (RF), Destek Vektör Makineleri (SVM), MLP (Yapay Sinir Ağı), XGBoost (Extreme Gradient Boosting), DNN (Deep Neural Networks),SNN(Spiking Neural Networks) modellerinin saniye cinsinden yürütme sürelerine bağlı olarak veri işleme hızları gösterilmektedir.

Tablo 3'de verilen süreler, modelin eğitim sürelerini temsil etmektedir. Eğitim süresi, modelin eğitim verisi üzerinde öğrenme sürecini tamamlaması için geçen toplam zamanı ifade eder. Bu süreler, her bir algoritmanın veri setini öğrenme kapasitesini ve verimliliğini karşılaştırmak için kullanılır.

Tablo 2. Oluşturulan modellerin başarım ölçütleri.

Table 2. Performance Metrics of the Developed Models.

Algoritma	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.662921	0.643363	0.662921	0.620993
Naive Bayes	0.619188	0.680176	0.619188	0.624862
J48	0.672083	0.655288	0.672083	0.648410
Random Forest	0.672083	0.655288	0.672083	0.648410
SVM	0.716742	0.701312	0.716742	0.658597
MLP	0.670268	0.657060	0.670268	0.622240
XGBoost	0.719651	0.701647	0.719651	0.670359
DNN	0.720709	0.700010	0.720709	0.680230
SNN	0.665079	0.647068	0.665099	0.635456

Tablo 3. Yürütme süreleri.

Table 3. Execution Times.

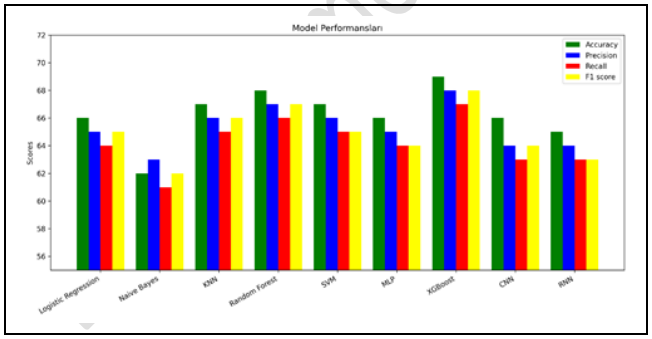
Model	Zaman(s)
Logistic Regression	0.016075
Naive Bayes	0.003987
J48	0.004983
Random Forest	0.283277
SVM	6.007320
MLP	1.592360
XGBoost	0.063785
DNN	11.930192
SNN	14.064498

Naive Bayes ve J48 gibi daha basit algoritmaların eğitim süreleri oldukça kısadır.

Random Forest ve XGBoost gibi daha karmaşık algoritmaların ise eğitim süreleri daha uzundur.

SVM, DNN ve SNN gibi daha karmaşık ve hesaplama açısından yoğun modellerin eğitim süreleri ise oldukça uzundur.

Şekil 13'te Modellere ait performanslar grafik olarak sunulmuştur.



Şekil 13. Model performans grafiği.

Figure 13. Model Performance Graph.

J48 algoritmasının başarı oranı yüksek bir şekilde sınıflandırma yapması nedeniyle, karar ağacı J48 algoritması tarafından oluşturulmuştur. Karar ağacının oluşturulması sürecinde, veri kümesinin bölünmesi için yaygın olarak kullanılan bir kriter entropidir. Entropi, bir veri kümesindeki belirsizliği ölçen bir kavramdır. Yani, bir veri kümesinin ne kadar homojen veya heterojen olduğunu değerlendirmek için kullanılır. Entropi hesaplamak için Denklem (5) kullanılır.

$$Entropi(T) = \sum_{t=1}^{n} p_i \log_2(p_i) \quad (5)$$

Entropi, her bir sınıfın veri kümesindeki dağılımına dayanarak hesaplanır. Eğer bir veri kümesi tamamen homojen bir sınıfa sahipse, yani tüm veriler aynı sınıfa aitse, entropi değeri düşüktür. Ancak, veri kümesi farklı sınıflara ait verileri içeriyorsa, entropi değeri yüksektir, yani veri kümesi daha heterojendir.

Entropi, karar ağacının dallarının oluşturulmasında ve veri kümesinin sınıflandırılmasında da kullanılır. Karar ağacı, veri kümesinin içerdiği entropi değerlerini minimize etmek için en iyi bölünme noktalarını seçerek verileri sınıflandırır [33].

Şekil 14'de J48 algoritmasına ait karar ağacı sunulmaktadır. Karar ağacı, Takibi Devam Eden Kredi kök düğümünden başlayarak dallanmıştır. Daha sonrasında Yas Skalası düğümü oluşturularak yaprak düğümlere ulaşılmıştır.

6.3 Sınıflandırma modellerinin ortalama başari karşilaştirması (10-Fold Cross Validation)

Yürütülen 10 katlı çapraz doğrulama analizleri sonucunda, optimize edilmiş hiperparametrelerle Derin Sinir Ağı (DNN) modeli en yüksek ortalama doğruluğa (%72,79) ulaşmıştır. Diğer modeller ise %69 ile %71.2 arasında değişen doğruluk oranları sergilemişlerdir.

Özellikle MLP, XGBoost ve SVM modelleri, benzer doğruluk performanslarıyla dikkat çekmekte olup, DNN modeline yakın sonuçlar elde edilmiştir. Geleneksel yöntemler olan Decision Tree ve Naive Bayes ise diğerlerine göre daha düşük başarı göstermiştir.

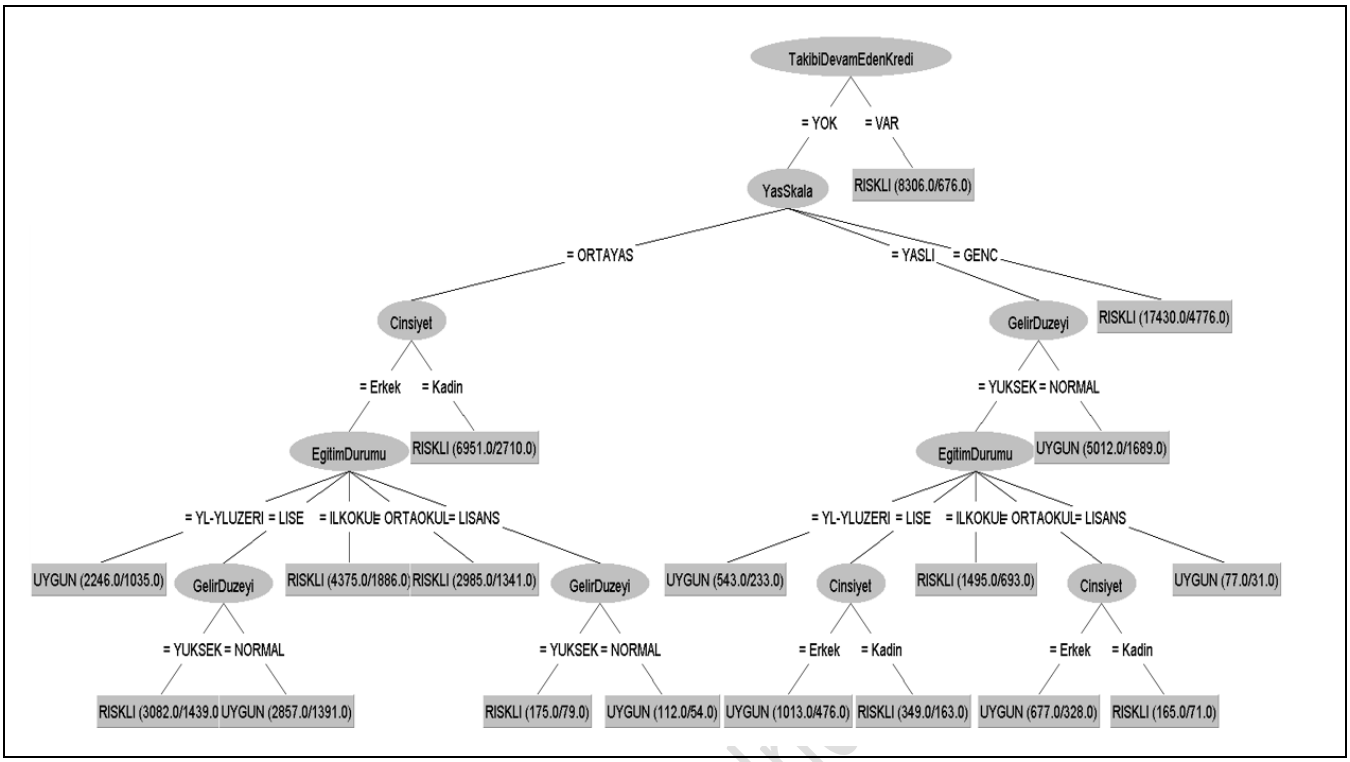
Bu sonuçlar, karmaşık verisetlerinde optimize edilmiş derin öğrenme yöntemlerinin daha iyi performans sunabileceğini göstermektedir. Bununla birlikte, uygulama alanı ve kaynak kısıtları doğrultusunda diğer modeller de tercih edilebilir. 10 katlı çapraz doğrulama sonuçları Tablo 4'de sunulmuştur.

Model performansları 10 katlı çapraz doğrulama (fold) kullanılarak değerlendirildi. Tablo 4'te her modelin fold bazında elde edilen ortalama doğruluk ve standart sapmaları verilmiştir.

Ortalama doğruluk açısından en iyi performans, %72.16 ile derin sinir ağı (DNN) modelinde gözlemlenmiştir. En düşük ortalama doğruluk ise %69.88 ile Naive Bayes modelindedir.

Şekil 15'de, 10 katlı çapraz doğrulama yöntemi kullanılarak elde edilen farklı modellerin fold bazında doğruluk performansları gösterilemektedir. Model doğrulukları fold'lar arasında genellikle tutarlı olup, DNN modeli diğer modellere kıyasla daha yüksek ve stabil bir doğruluk sergilemektedir. Bu grafik, modellerin genel performansını ve sonuçların varyansını görsel olarak değerlendirmeye olanak sağlamaktadır.

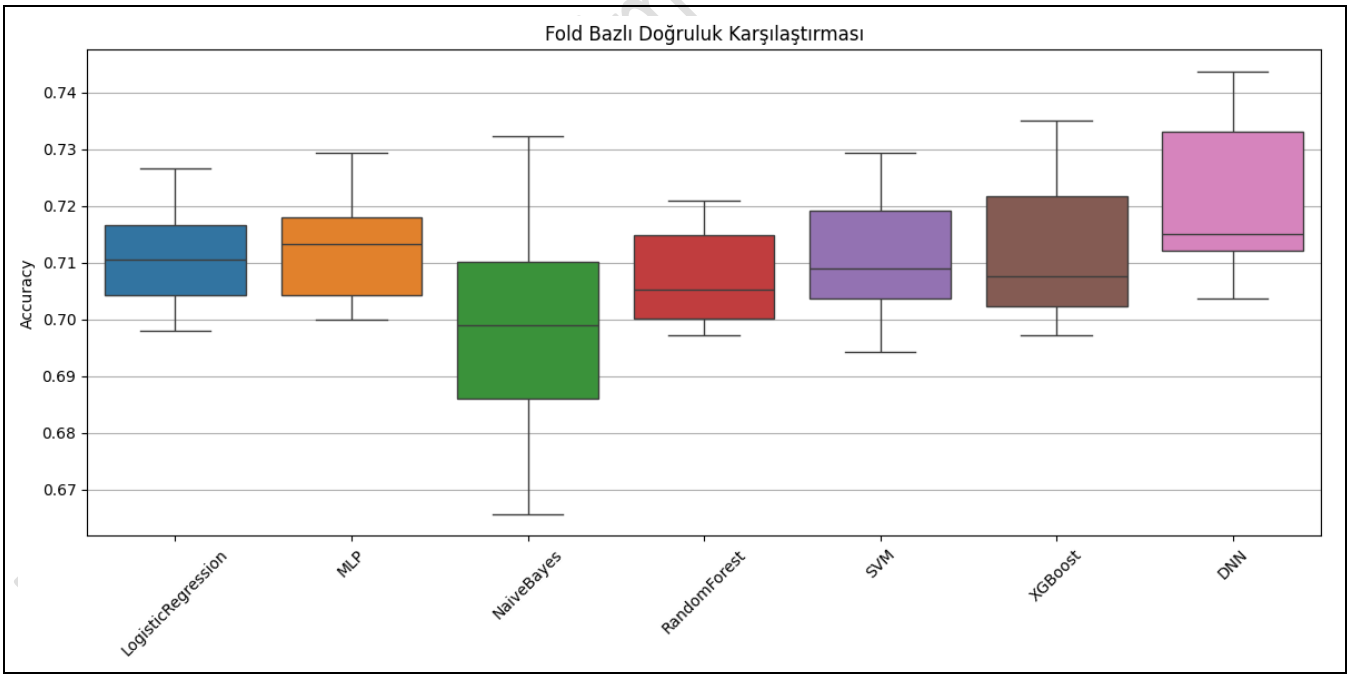
İstatistiksel anlamlılığı test etmek için Logistic Regression ile diğer modeller arasında eşleştirilmiş (paired) Wilcoxon işaret testi ve t-testi uygulanmıştır. Sonuçlar Tablo 5'te yorumlanmıştır.



Şekil 14. J48 algoritması karar ağacı.

Figure 14. Decision tree of the J48 algorithm.

Tablo 4. Katlı Çapraz Doğrulama Sonuçları: Ortalama Doğruluk, Standart Sapma ve En İyi Parametreler.



Şekil 15. Fold Bazlı Doğruluk Performanları.

Figure 15. Fold-Based Accuracy Performances.

Table 4. K-Fold Cross-Validation Results: Mean Accuracy, Standard Deviation, and Best Parameters

Algoritma	Means Accuracy	Std Dev	Best Params
Logistic Regression	0.7105	0.0161	{'C': 0.1, 'penalty': 'l2', 'solver': 'liblinear'}
Naive Bayes	0.6988	0.0222	{'var_smoothing': 1e-09}
Random Forest	0.7074	0.0191	{'criterion': 'gini', 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 200}
SVM	0.7111	0.0150	{'C': 10, 'gamma': 'scale', 'kernel': 'rbf'}
MLP	0.7122	0.0146	{'activation': 'relu', 'alpha': 0.0001, 'hidden_layer_sizes': (50, 50), 'solver': 'sgd'}
XGBoost	0.7122	0.0163	{'colsample_bytree': 0.6, 'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100, 'subsample': 0.8}
DNN	0.7279	0.0152	Keras Tuner ile optimize edilmiş

Tablo 5. DNN Modeli ile Diğer Modellerin İstatistiksel Karşılaştırması (Wilcoxon ve Paired t-Test).

Table 5. Statistical Comparison of the DNN Model with Other Models (Wilcoxon and Paired t-Test).

Karşılaştırma	Wilcoxon İstatistiği	Wilcoxon p-değeri	Paired t-Test t-değeri	Paired t-Test p-değeri	Yorum
DNN vs LogisticRegression	2.0000	0.0439	2.5844	0.0295	Her iki testte de anlamlı fark var. DNN, Logistic Regression'dan daha iyi.
DNN vs MLP	3.0000	0.1090	1.7261	0.1162	Anlamlı fark yok, performanslar benzer.
DNN vs NaiveBayes	0.0000	0.0078	5.6101	0.0002	DNN, Naive Bayes'e kıyasla anlamlı şekilde üstün.
DNN vs RandomForest	2.0000	0.0439	2.7882	0.0206	DNN, Random Forest'dan anlamlı olarak daha iyi.
DNN vs SVM	2.0000	0.0439	2.0981	0.0679	Wilcoxon anlamlı, t-testi anlamlılık sınırında. DNN muhtemelen daha iyi.
DNN vs XGBoost	3.0000	0.1090	1.6261	0.1353	Anlamlı fark yok, performanslar birbirine yakın.

İstatistiksel anlamlılık testi kapsamında, DNN modeli ile diğer sınıflandırma modellerinin performansları Wilcoxon işaret testi ve eşleştirilmiş (paired) t-testi ile karşılaştırılmıştır. DNN ile Logistic Regression, Naive Bayes ve Random Forest modelleri arasındaki farklar her iki testte de istatistiksel olarak anlamlı bulunmuştur; bu durum DNN modelinin bu modellerden üstün performans sergilediğini göstermektedir. Buna karşın, DNN ile MLP ve XGBoost modelleri arasında anlamlı bir fark gözlenmemiştir; dolayısıyla bu modellerin performansları birbirine yakın kabul edilebilir. DNN ve SVM karşılaştırmasında ise Wilcoxon testi anlamlılık gösterirken, paired t-testi anlamlılık sınırında ($p \approx 0.068$) sonuç vermiştir. Bu durum, DNN'nin SVM'ye göre daha iyi performans sergileme olasılığının yüksek olduğunu ancak bu farkın daha güçlü istatistiksel kanıtlarla desteklenmesi gerektiğini ortaya koymaktadır. Genel olarak, DNN modeli çoğu rakibine kıyasla istatistiksel olarak üstün performans gösterirken, bazı modellerle benzer başarım sergilemektedir. Bu sonuçlar, kredi skoru tahmininde DNN'nin güçlü bir alternatif olduğunu desteklemektedir.

7 Birliktelik analizi (Association rules)

Kredi skoru tahminine yönelik veri analizi sürecinin bir sonraki aşamasında birliktelik analizi uygulanmıştır. Bu bölümde, Apriori algoritmasının uygulanma yöntemi ve elde edilen sonuçlar detaylı şekilde sunulacaktır.

Birliktelik analizi, iki temel metrik kullanarak kuralları değerlendirir: destek (support) ve güven (confidence). Destek, birliktelik kuralındaki öğelerin birlikte ne sıklıkla görüldüğünü ifade ederken, güven, birliktelik kuralının ne kadar doğru olduğunu belirtir [36].

Birliktelik analizi, veri setindeki öğeler arasındaki sık görülen ilişkileri keşfetmek amacıyla kullanılan bir veri madenciliği tekniğidir. Genellikle büyük veri kümelerinde veya karmaşık veri tablolarında, öğeler arasındaki anlamlı bağlantıların ortaya çıkarılmasında tercih edilmektedir [37].

Analizde, birliktelik kurallarının değerlendirilmesinde iki temel metrik kullanılmaktadır: destek (support) ve güven (confidence). Destek, belirli bir öğe setinin veri kümesinde ne sıklıkla birlikte ortaya çıktığını ifade ederken; güven ise, kuralın doğruluğunu yani koşullu olasılığını belirtmektedir [37].

7.1 Birliktelik analizi ile ilgili tanımlar

Kredi skoru analizinde kullanılan tüm öğelerin kümesi, I = {i1, i2, i3, ..., id}, ve bu analizdeki bütün işlemlerin kümesi T = {t1, t2, t3, ..., tN} olarak tanımlansın. Her bir ti işlemi, I öğe kümesinin bir alt kümesini içermektedir. Birliktelik analizi, öğe kümesini temsil eden bir veya daha fazla öğeyi içeren kümeyi kullanır.

Birliktelik Kuralı: $X \rightarrow Y$ iken $X \subset I$, $Y \subset I$, ve $X \cap Y = \emptyset$

Öğe Kümesi en az bir öğeyi içeren bir koleksiyondur.

k-öğe kümesi içerisinde k adet öğe içeren koleksiyondur.

Destek sayısı (σ) öğe kümesinin görülme sıklığını ifade eder.

Destek (Support) $X \rightarrow Y$ kuralındaki X ve Y öğelerini içeren işlemlerin, tüm işlemlere oranıdır.

Güven (Confidence) $X \rightarrow Y$ kuralında Y öğe kümesindeki elemanların X öğe kümesi elemanlarının bulunduğu işlemlerdeki sıklığını göstermektedir.

Association Rules, veri madenciliği alanında büyük bir öneme sahip olan bir analiz tekniğidir. Bu teknik, veri kümelerindeki öğeler arasındaki ilişkileri ve desenleri belirlemek için kullanılır. Association Rules, birçok uygulama alanında kullanılır ve işletmelere büyük fayda sağlar [37].

Association Rules'ün kullanım alanlarından biri, perakende sektörüdür. Mağazalar, müşterilerin alışveriş sepetlerini analiz ederek hangi ürünlerin bir arada satın alındığını

belirleyebilirler. Örneğin, kahve alan müşterilerin sıklıkla krema da aldığını gözlemleyebilirler. Bu tür ilişkileri belirlemek, ürün yerleştirme stratejilerini ve promosyonları optimize etmeye yardımcı olur.

Association Rules ayrıca web sitesi kişiselleştirme için de kullanılır. E-ticaret siteleri, kullanıcıların geçmiş tarama ve alışveriş davranışlarına dayalı olarak öneriler sunar. Örneğin, bir kullanıcı belirli bir ürünü görüntülerse, web sitesi aynı kategorideki diğer ürünleri önerir. Bu, müşteri deneyimini iyileştirmeye yardımcı olur[38].

Sağlık sektörü de Association Rules'ü kullanır. Hastaların semptomlarını ve tıbbi geçmişlerini analiz ederek hangi testlere veya tedavi yöntemlerine ihtiyaç duyduklarını belirleyebilirler. Bu, doğru teşhislerin yapılmasına ve tedavi seçeneklerinin optimize edilmesine yardımcı olabilir.

Association Rules'ün önemi, işletmelerin veriye dayalı kararlar almasını desteklemesinde yatar. Bu kural analizi, pazarlama stratejilerini optimize etmek, envanter yönetimini iyileştirmek ve iş süreçlerini verimli hale getirmek için kullanılabilir. Ayrıca, büyük veri kümelerinde gizli ilişkileri ve desenleri keşfetmek için kullanılabilir, bu da yeni fırsatları tanımlamak için büyük bir potansiyel sunar [39].

Association Rules, veri madenciliği alanında güçlü bir araçtır ve birçok uygulama alanında büyük bir öneme sahiptir. Bu teknik, işletmelere daha iyi kararlar almaları ve müşteri deneyimini iyileştirmeleri konusunda yardımcı olur.

7.2 Birliktelik analizi kurallarının değerlendirilmesinde kullanılan metrikler

conf (confidence): Bu metrik, birliktelik kuralının doğruluk oranını ifade eder. Conf değeri 0 ile 1 arasında bir değer alır ve 1'e yaklaştıkça kuralın daha doğru olduğunu gösterir. Örneğin, conf değeri 0.98 ise, kuralın %98 doğruluk oranına sahip olduğunu ifade eder.

$$Confidence\ (X \rightarrow Y) = \frac{support(A,B)}{support(A)} \quad (6)$$

lift: Lift, birliktelik kuralının beklentiden ne kadar daha iyi performans gösterdiğini gösteren bir metriktir. Lift değeri 1'den büyükse, X ve Y öğeleri arasındaki ilişkinin diğer rastgele bir ilişkiye kıyasla daha güçlü olduğunu gösterir. Örneğin, lift değeri 1.14 ise, kuralın diğer rastgele bir ilişkiden %14 daha güçlü olduğunu ifade eder.

$$Lift\ (A \rightarrow B\ ) = sup(A,B) - (sup(A) + sup(B)) \quad (7)$$

lev (leverage): Leverage, X ve Y öğeleri arasındaki gerçekleşme oranının beklentiden ne kadar farklı olduğunu ölçen bir metriktir. Lev değeri -1 ile 1 arasında bir değer alır. Lev değeri 0'a yaklaştıkça, X ve Y öğeleri arasındaki ilişki beklentilere daha yakın olduğunu gösterir. Örneğin, lev değeri 0.01 ise, ilişkinin beklentilere yakın olduğunu ifade eder.

$$Leverage\ (A \rightarrow B) = sup(A,B) - (sup(A) * sup(B)) \quad (8)$$

conv (conviction): Conviction, birliktelik kuralının yanlışlık oranını ifade eder. Conv değeri 1'den büyükse, kuralın yanlışlık oranının yüksek olduğunu ve kuralın daha güvenilir olmadığını gösterir. Örneğin, conv değeri 7.08 ise, kuralın yanlışlık oranının diğer rastgele bir ilişkiye kıyasla 7.08 kat daha yüksek olduğunu ifade eder [39].

$$Conviction\ (A \rightarrow B) = \frac{1 - support(A,B)}{1 - Confidence(A \rightarrow B)} \quad (9)$$

7.3 Apriori algoritması

Apriori algoritması, en sık kullanılan ve bilinen birliktelik kuralı çıkarım algoritmasıdır. Apriori'ye göre, "Eğer bir k-öğe kümesi minimum destek kriterini sağlıyorsa, o kümenin herhangi bir alt kümesi de minimum destek kriterini sağlar." şeklindedir. Bir öğe kümesi, birden fazla öğeden oluşan bir kümeyi ifade eder. k-öğe kümesi (k-itemset) ise içinde k adet öğe bulunan bir kümedir. Apriori, (k+1) sık geçen öğe kümesini bulmak için en sık geçen öğe kümesine dayanır. Sık geçen öğe kümelerini tespit etmek için minimum destek kriterini sağlayan başlangıçtaki sık geçen öğe kümesi belirlenir ve bu işlem, algoritma sık geçen öğe kümeleri bulunamayıncaya kadar tekrarlanır [37].

Apriori algoritması ile gerçekleştirilen birliktelik analizi için minimum destek değeri %10 ve minimum güven değeri %90 olarak belirlenmiştir. Apriori algoritması kullanılarak gerçekleştirilen birliktelik analizi sonucunda 18 adet güven değeri yüksek kural oluşturulmuştur. Oluşturulan kurallar aşağıdaki gibidir.

Erkeklerin yüksek gelir düzeyine ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir. Bu kural, kredi skor tahmini üzerinde önemli bir ilişkiyi ifade etmektedir.

Yüksek gelir düzeyine ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.

Lise eğitim düzeyine ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.

Normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.

- Cinsiyeti erkek ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Yaş aralığı orta yaş ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Cinsiyeti erkek ve yaş aralığı orta yaş ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Normal gelir düzeyine ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Cinsiyeti erkek ve normal gelir düzeyine ve normal kredi skoruna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Yüksek gelir düzeyine ve genç yaş aralığına sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Cinsiyeti erkek ve takibi devam eden bir krediye sahipse kredi skoru Riskli olduğunu göstermektedir.
- Takibi devam eden bir krediye sahipse kredi skoru Riskli olduğunu göstermektedir.
- Lise eğitim durumuna ve Yüksek gelir düzeyine sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.

- Yüksek gelir düzeyine sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Yüksek gelir düzeyine ve orta yaş aralığına sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Cinsiyeti erkek ise ve yüksek gelir düzeyine sahip ve lise eğitim durumuna sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Cinsiyeti erkek ise yüksek gelir düzeyine ve orta yaş aralığına sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.
- Cinsiyeti erkek ise yüksek gelir düzeyine sahip olduğu durumlarda takibi devam eden bir kredilerinin olmadığını göstermektedir.

Tablo 6'da Apriori algoritması ile elde edilen birliktelik kurallarını ve bu kuralların ölçümleri gösterilmiştir.

Destek (Support) değeri, kuralın veri setinde ne sıklıkla ortaya çıktığını belirtmektedir. Güven (Confidence) değeri, kuralın doğruluk oranını ifade ederken, kaldırma (Lift) değeri, kuralın tesadüfi oluşumdan ne kadar bağımsız olduğunu göstermektedir.

Örneğin, "Gelir düzeyi yüksek → Takibi devam eden kredi yok" kuralının destek, güven ve kaldırma değerleri sırasıyla 0.370, 0.907 ve 1.060 olarak bulunmuştur. Bu değerler, yüksek gelir düzeyine sahip bireylerin takibi devam eden kredilerinin olmaması durumunun veri setinde sık ve güvenilir bir şekilde gözlendiğini göstermektedir. Benzer şekilde, "Takibi devam eden kredi var → Kredi skoru riskli" kuralı destek, güven ve kaldırma değerleri ile birlikte, takibi devam eden kredisi olan bireylerde kredi skorunun riskli olma olasılığının yüksek olduğunu ortaya koymaktadır.

Diğer kurallar da kredi skoru, cinsiyet, yaş aralığı ve eğitim durumu gibi faktörler ile takibi devam eden kredi durumları arasındaki ilişkileri detaylı bir şekilde göstermektedir. Bu kurallar, veri seti üzerindeki ilişkileri ve olası kredi skoru tahminlerini anlamada önemli bilgiler sağlamaktadır.

8 Sonuçlar

Veri madenciliği alanında, bilgiye erişme yöntemleri ve kredi risk tahmini gibi önemli uygulama alanlarında birçok algoritma bulunmaktadır. Bu algoritmaların hangisinin daha üstün olduğu konusu üzerine birçok çalışma yapılmış, ancak bu çalışmalardan farklı sonuçlar elde edilmiştir. Bu çeşitlilik, işlem başarısının kullanılan veri kaynağına, veri ön işlemeye ve algoritma parametrelerine bağlı olduğunu göstermektedir. Bu çalışma, finans kuruluşlarına ait verileri kullanarak, farklı sınıflandırma algoritmalarının kredi riski tahmini konusundaki etkinliğini incelemeyi amaçlamıştır.

Bu çalışmada, Python programlama dilinde yaygın olarak kullanılan scikit-learn, TensorFlow, ve Keras gibi kütüphanelerden faydalanılarak modellerin oluşturulması ve eğitilmesi gerçekleştirilmiştir. Bu kütüphaneler, hem makine öğrenimi hem de derin öğrenme modellerinin hızlı bir şekilde prototipini oluşturmayı ve optimize etmeyi sağlamaktadır. Çalışmalar, yüksek performanslı bir Intel Core i7 işlemci ve 32 GB RAM ile donatılmış bir bilgisayar ortamında yürütülmüştür.

J48 ve RandomForest gibi karar ağaçları tabanlı modeller, hızlı eğitim süreleri ve kabul edilebilir performanslarıyla dikkat çekmiştir. Ancak SVM gibi bazı modeller, yüksek doğruluk sağlasa da eğitim süreleri daha uzun olabileceğinden pratik kullanımlarda dezavantajlar doğurabilir. Her bir modelin avantajları ve dezavantajları dikkate alınarak, uygulama senaryosuna en uygun modelin seçilmesi kritik öneme sahiptir.

En yüksek doğruluğa ve F1-score değerlerine sahip model, DNN (Deep Neural Network) olarak belirlenmiştir. Bu model, özellikle kredi skoru tahmini gibi uygulamalarda diğerlerine göre üstün performans sergilemiştir. DNN algoritması, accuracy, recall ve F1-Score metriklerinde diğer algoritmaların önünde yer almakta ve daha dengeli bir performans sergilemektedir. Ayrıca, bu modellerin eğitim sürelerinin diğerlerine göre daha kısa olması, gerçek zamanlı veya hızlı sonuç gerektiren senaryolarda avantaj sağlamaktadır.

Bu çalışma, kredi skor tahmininde çeşitli makine öğrenimi algoritmalarının karşılaştırmalı bir analizini sunmuş ve DNN'nin üstün performansını ortaya koymuştur. Bununla birlikte, her finansal kurumun iş süreçleri ve veri gereksinimlerinin farklılık gösterdiği dikkate alındığında, model seçimi ve uygulama adımlarının her kuruma özgü şekilde özelleştirilmesi gerekmektedir. Gelecekteki araştırmalar, istatistiksel güvenilirliği artırmaya yönelik yöntemlerin kullanılması ve farklı zaman dilimlerine ait verilerin değerlendirilmesiyle model sonuçlarının doğruluğunu daha da güçlendirebilir.Farklı makine öğrenimi tekniklerinin karşılaştırmalı analizini sunarak, veri bilimi uygulayıcılarına çeşitli senaryolarda en etkili çözümü seçme konusunda rehberlik etmeyi amaçlamaktadır.

Tablo 6. Apriori algoritması ile elde edilen birliktelik kuralları ve ölçümleri

Table 6. Association Rules and Metrics Obtained Using the Apriori Algorithm

Antecedents	Consequents	Support	Confidence	Lift
GelirDuzeyi_YUKSEK	TakibiDevamEdenKredi_YOK	0.3702	0.9074	1.0595
TakibiDevamEdenKredi_VAR	KrediSkoru_RISKLI	0.1318	0.9186	1.4473
KrediSkoru_NORMAL	TakibiDevamEdenKredi_YOK	0.3536	0.9680	1.1303
Cinsiyet_Erkek, GelirDuzeyi_YUKSEK	TakibiDevamEdenKredi_YOK	0.2766	0.9046	1.0562
Cinsiyet_Erkek, TakibiDevamEdenKredi_VAR	KrediSkoru_RISKLI	0.1005	0.9190	1.4480
Cinsiyet_Erkek, KrediSkoru_NORMAL	TakibiDevamEdenKredi_YOK	0.2559	0.9665	1.1286
GelirDuzeyi_YUKSEK, EgitimDurumu_LISE	TakibiDevamEdenKredi_YOK	0.1409	0.9086	1.0609
EgitimDurumu_LISE, KrediSkoru_NORMAL	TakibiDevamEdenKredi_YOK	0.1290	0.9689	1.1314
GelirDuzeyi_NORMAL, KrediSkoru_NORMAL	TakibiDevamEdenKredi_YOK	0.2020	0.9602	1.1212
GelirDuzeyi_YUKSEK, YasSkala_GENC	TakibiDevamEdenKredi_YOK	0.1102	0.9224	1.0771
GelirDuzeyi_YUKSEK, YasSkala_ORTAYAS	TakibiDevamEdenKredi_YOK	0.1852	0.9073	1.0594
GelirDuzeyi_YUKSEK, KrediSkoru_NORMAL	TakibiDevamEdenKredi_YOK	0.1516	0.9785	1.1426
KrediSkoru_NORMAL, YasSkala_ORTAYAS	TakibiDevamEdenKredi_YOK	0.1761	0.9656	1.1274

Gelecekteki çalışmalar için önerilerimiz şunlardır:

- İstatistiksel güvenilirliği artırmak amacıyla p-değerleri, çapraz doğrulama ve bootstrapping gibi yöntemlerin sistematik kullanımı,
- Model sonuçlarının farklı ekonomik koşullar ve piyasa dalgalanmaları altında test edilmesi,
- Veri ön işleme ve özellik mühendisliği süreçlerinin optimize edilerek modellerin farklı veri setlerine uyarlanması,
- Finansal kurumların ihtiyaçlarına göre model özelleştirme yöntemlerinin geliştirilmesi,
- Birliktelik analizi ve diğer veri madenciliği tekniklerinin kredi risk tahminine entegrasyonu ve performans karşılaştırmaları.
- Bu kapsamlı değerlendirme ve öneriler, kredi riski tahmini alanındaki araştırmacılar için rehber niteliğinde olup, daha güvenilir ve uygulanabilir modeller geliştirilmesine katkı sağlayacaktır.

Bu çalışma, istatistiksel anlamlılığın daha iyi anlaşılması gerektiğini vurgulamaktadır. İleriki çalışmalarda, p-değerleri veya çapraz doğrulama gibi istatistiksel yöntemlerin kullanılması sonuçların güvenilirliğini daha fazla artırabilir. Kredi riski tahmini, ekonomik dönemlere ve piyasa koşullarına bağlı olarak değişebilir ve bu çalışma, belirli bir döneme ait verilerle sınırlıdır. Gelecekteki çalışmalarda farklı dönemlere ait verilerin de incelenmesi, daha geniş bir bakış açısı sunabilir.

Ayrıca, finans kuruluşlarının boyutları, iş süreçleri ve veri gereksinimleri farklılık gösterebilir. Bu nedenle, bu çalışmanın sonuçları ve önerileri, her bir kuruluşun ihtiyaçlarına özgü olarak uyarlanmalıdır. Finans kuruluşlarının kredi verme süreçlerinin farklılıkları dikkate alındığında, bu çalışmanın sonuçları, her bir kuruluşun kendi özgün gereksinimlerine göre şekillendirilmelidir.

Sonuç olarak, bu çalışma, kredi skor tahmininde çeşitli makine öğrenimi algoritmalarının karşılaştırmalı bir analizini sunmuş ve DNN'nin üstün performansını ortaya koymuştur. Bununla birlikte, her finansal kurumun iş süreçleri ve veri gereksinimlerinin farklılık gösterdiği dikkate alındığında, model seçimi ve uygulama adımlarının her kuruma özgü şekilde özelleştirilmesi gerekmektedir. Gelecekteki araştırmalar, istatistiksel güvenilirliği artırmaya yönelik yöntemlerin kullanılması ve farklı zaman dilimlerine ait verilerin değerlendirilmesiyle model sonuçlarının doğruluğunu daha da güçlendirebilir.

9 Kaynaklar

[1] Han J., Pei J., Kamber M. *Data Mining: Concepts and Techniques*. Elsevier, 2012.

[2] Tan P.N., Steinbach M., Kumar V. *Introduction to Data Mining*. Pearson Education India, New Delhi, 2016.

[3] Provost F., Fawcett T. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media, 2013.

[4] Noman S.M., Fadel Y.M., Henedak M.T., Attia N.A., Essam M., Elmaasarawii S., Fouad F.A., Eltasawi E.G., Al-Atabany W. "Leveraging survival analysis and machine learning for accurate prediction of breast cancer recurrence and metastasis". *Scientific Reports*, 15, 3728, 2025. https://doi.org/10.1038/s41598-025-87622-3

[5] Duda R.O., Hart P.E., Stork D.G. *Pattern Classification* (2nd ed.). Wiley, 2012.

[6] Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning*. Springer, 2009.

[7] Rahman F.M.S., Islam M.S. "Data clustering: Application and trends". *Artificial Intelligence Review*, 2022.

[8] James G., Witten D., Hastie T., Tibshirani R. *An Introduction to Statistical Learning* (2nd ed.). Springer, 2021.

[9] Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016.

[10] Shi S., Tse R., Luo W., D'Addona S., Pau G. "Machine learning-driven credit risk: A systemic review". *Neural Computing and Applications*, 34(17), 14327–14339, 2022.

[11] Moreira E., Moro S., Cortez P. "Deep learning and machine learning techniques for credit scoring: A review". *Communications in Computer and Information Science*, 2069, 30–61, 2024. https://doi.org/10.1007/978-3-031-57639-3_2

[12] Brown J., Smith A., Lee M. "Comparison of decision trees and support vector machines for credit scoring". *Journal of Financial Data Science*, 3(2), 45–60, 2021.

[13] Chen L., Zhang Y. "Application of XGBoost in large-scale credit scoring". *International Journal of Machine Learning*, 9(1), 78–89, 2022.

[14] Li H., Wang Q., Zhao J. "Neural networks for credit risk assessment: Accuracy and training time". *Neural Computing and Applications*, 32(5), 1423–1435, 2020.

[15] Zhou F., Huang T., Yang P. "Deep learning models for credit risk prediction: A comparative study". *Expert Systems with Applications*, 215, 119526, 2023.

[16] Ghosh R., Kumar S., Roy S. "Logistic regression vs. other classifiers in credit scoring". *Journal of Risk and Financial Management*, 13(7), 157, 2020.

[17] Zontul M., Polat H., Yıldırım T. "Customer credit rating estimation using machine learning methods". *Applied Soft Computing*, 106, 107312, 2021.

[18] Arram A., Ayob M., Albadr M.A.A., Sulaiman A., Albashish D. "Comparing the effectiveness of machine learning and deep learning models in student credit scoring: A case study in Vietnam". *Applied Soft Computing*, 142, 110839, 2023.

[19] Hussain A., Khan M.A., Ahamed A.K.S., Kousarziya S. "Enhancing credit scoring models with artificial intelligence: A comparative study of traditional methods and AI-powered techniques". *Expert Systems with Applications*, 230, 120099, 2023.

[20] Mhlanga D. "Machine learning approaches to credit risk: Comparative study of Islamic and conventional banks". *Journal of Risk and Financial Management*, 14(12), 608, 2021.

[21] Sarıman G. "Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: K-Means ve K-Medoids kümeleme algoritmalarının karşılaştırılması". *Süleyman Demirel Üniversitesi Fen Bilimleri Dergisi*, 2011.

[22] Matei G. "A collaborative approach of Business Intelligence systems". *Journal of Applied Collaborative Systems*, 2(2), 91–101, 2010.

[23] Thomas L.C. *Consumer Credit Models: Pricing, Profit and Portfolios* (2nd ed.). Oxford University Press, 2020.

[24] Hair J.F., Black W.C., Babin B.J., Anderson R.E., Tatham R.L. "Multivariate Data Analysis". 6. baskı. Pearson Prentice Hall, Upper Saddle River, NJ, 2006.

[25] Mitra B., Diaz F., Craswell N. *Introduction to Information Retrieval* (2nd ed.). Cambridge University Press, 2018.

[26] Rokach L., Maimon O. *Data Mining with Decision Trees: Theory and Applications*. World Scientific, 2014.
[27] Breiman L. "Random forests". *Machine Learning*, 45, 5–32, 2001.
[28] Huang C.L., Wang C.J. "A GA-based feature selection and parameters optimization for support vector machines". *Expert Systems with Applications*, 31(2), 231–240, 2006.
[29] Markham K. "Simple guide to confusion matrix terminology". http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/, 2014.
[30] Doğan E. "The Effect of Innovation on Competitiveness". *Istanbul University Econometrics and Statistics e-Journal*, No. 24, 60–81, 2016.
[31] Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016.
[32] Chen T., Guestrin C. "XGBoost: A scalable tree boosting system". *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, 2016.
[33] Ruder S. "An overview of gradient descent optimization algorithms". *arXiv preprint arXiv:1609.04747*, 2016.
[34] Diehl P.U., Cook M. "Unsupervised learning of digit recognition using spike-timing-dependent plasticity". *Frontiers in Computational Neuroscience*, 9, 99, 2015.
[35] Powers D.M.W. "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation". *Journal of Machine Learning Technologies*, 2(1), 37–63, 2011.
[36] Agrawal R., Imieliński T., Swami A. "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207–216, 1993.
[37] Zaki M.J., Meira W. *Data Mining and Analysis: Fundamental Concepts and Algorithms* (2nd ed.). Cambridge University Press, 2020.
[38] Sahami Shirazi A., Norouzi F. "A review on association rule mining methods and applications". *Journal of Big Data*, 8(1), 86, 2021.
[39] Kotsiantis S.B., Zaharakis I., Pintelas P. "Supervised machine learning: A review of classification techniques". *Emerging Artificial Intelligence Applications in Computer Engineering*, 159–184, 2020.