Bölüm 2

XGboost Algoritması

Gökhan Korkmaz[1]

Özet

Aşırı Gradyan Artırma Algoritması, kısa adıyla "XGBoost" (Extreme Gradient Boosting Algorithm), Karar Ağaçlarının (KA) özelleştirilmiş bir formu olup sınıflandırma, tahmin ve sıralama yöntemi olarak literatürde ön plana çıkmaktadır.

2015 Bilgi Keşfi ve Veri Madenciliği (Knowledge Discovery and Data Mining-KDD) kupasında seçilen en iyi 10 çözümün tamamının kullanmış olduğu, XGBoost algoritması, tüm branşlardan araştırmacıların ilgisini çekmekte, verdiği sonuçlar itibariyle çok etkili, bir o kadar karmaşık ve popüler bir yöntemdir. Denetimli öğrenen ve düzenleme (regülasyon) faktörüyle aşırı uyumdan kaçınan ve makinenin CPU çekirdeğinin uygun şekilde kullanılmasıyla daha fazla hız ve performansa yol açan çok iş parçacıklı bir yaklaşım benimseyen XGBoost algoritması tüm sektörler için potansiyel vaad etmekte ve gelişime açık yapısıyla da araştırmacılar için önemli bir çalışma sahası teşkil etmektedir.

Bu çalışmada XGBoost algoritmasının bugüne kadar ki evrimi, literatürdeki uygulama tecrübesi (yapılan çalışmalar – literatür taraması), içeriği, yapısı, işleyişi, parametreleri, Gradyan Artırma Algoritmasından (GAA) farkı, avantajları ve dezavantajları incelenmiş ve yöntem hakkında, ilgi duyan araştırmacılara, genel bir fikir sağlamak amaçlanmıştır.

1. GİRİŞ

İş hayatında olsun hayatın başka alanlarında olsun, işleri yürütme zorunluluğu bulunan insanların sık sık yerine getirmek zorunda kaldıkları ve hepsi sürecin sonunda adeta bir karın ağrısına dönen hayatî bir süreç var ki, bu sürecin sağlıklı yürütülmesi işlerin devamlılığı açısından çok hassas bir önem taşımaktadır. Bu süreç, "karar verme" faaliyetidir.

1 Dr. Öğr. Üyesi, Şırnak Üniversitesi, gokhankorkmaz@istanbul.edu.tr, ORCID: 0000-0002-1702-2965.

https://doi.org/10.58830/ozgur.pub574.c2353

Yöneticilerin en önemli vasfı olan sağlıklı karar verme süreci birçok anlamda iyi bir geleceğin anahtarını bünyesinde barındırmaktadır. Peki iş hayatında kararlar nasıl verilmektedir? Bu karar sürecini kısaca özetlemek gerekirse, öncelikle mevcut durum, doğru argümanlarla, doğru bir şekilde *algılanmaya* çalışılır. Burada "durumsallık analizi" diye başlıklandırabileceğimiz, "her durum kendine özgüdür ve verilmesi gereken cevap da mevut duruma özgün olmalıdır" teorisi çerçevesinde her durum için geçmiş tecrübelerden evet yararlanılabilir fakat her durumun kendine has parametreleri olduğu gerçeği göz ardı edilmemelidir. Algılanan duruma yakın bir tecrübe geçmişte yaşanmış ise geçmiş çözüm tecrübeleri masaya yatırılır. Mevcut durumun farkları sisteme entegre edilir. Uzman kimselerin sezgileri, yani bir genel *muhakeme,* bir değişken olarak, çözüm sürecine enjekte edilir ve tüm bu parametreler eşliğinde optimal *karar* oluşturulmaya çalışılır. Aslında algılamayla başlayan, muhakemeyle devam eden ve kararla sonuçlanan bu süreç, insanda, *zekâ* olarak tanımlanır.

Geçmişte yaşanan emsaller konuyu aydınlatıcı bir etki yapabilir fakat söz konusu olan binlerce pozitif ya da negatif diyabet vakasının yaşandığı bir hastane kayıtlarıysa, ya da binlerce kişinin kredi kullandığı ya da kapısından eli boş döndüğü bir banka kredi birimiyse, yani özetle geçmişte yaşanan tecrübe sayısı binlerle ifade ediliyorsa, tüm bu tecrübeyi süzecek veri madenciliği uygulamalarına şiddetle ihtiyaç duyulmaktadır.

Telekominikasyon, bilişim ve iletişim teknolojilerinin son derece yaygınlaşması neticesinde insanların da sordukları sorulara hızlı bir şekilde cevap beklemeleri bu karar süreçlerini otomatikleştirmeyi zorunlu kılmaktadır. Örneğin ödemesi olan bir gerçek ya da tüzel kişi, yapacağı kredi başvuruna birkaç dakika içinde geri dönüş beklemektedir. Bu durum bu tarz karar birimlerinde istihdam edilecek uzman yardımcılarının yanı sıra süreci hızlandıracak yazılımsal desteğe ve bu yazılımların içini dolduracak model derinliğine de ihtiyaç duymakta hatta bu süreci zaman zaman tamamen yazılımsal altyapıya terk etme güdüsü taşımaktadırlar. Yani eskiden insanların yaptıkları işi önce kontrollü olarak, sonrasında ise tamamen otomatik süreçlere terk etme günümüz iş süreçlerinin ruhunu oluşturmaktadır. Yani eskiden doğal zekâ ile çözülen bu problemlere bugün yapay zekâ ile alternatif aranmaktadır. Bu da yazılımlara gömülen ileri, karmaşık, model tabanlı, matematik destekli, veriden öğrenen ve durumsallık analizini göz önünde bulundurabilme kabiliyeti olan yapay zekâ yöntemleriyle mümkün olabilmektedir.

İnsanın yaptığı bir işi, bir makineye yaptırmak kolay değildir. Çünkü insanın her eyleminin arkasında, kapsamlı bir algılama sürecinin ardından

çok derin bir muhakeme ve nihayetinde karar verme ve bu kararı uygulama yetisi yatar.

İnsana ait olan süreçleri taklit eden yapay zekâya dönük çalışmalar aslında çok yeni değildir fakat özellikle son birkaç on yılda bu alanda çok önemli gelişmeler kaydedildi. Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM), Genetik Algoritma (GA), K-En Yakın Komşu (KEYK), Karar Ağaçları (KA) ve bunun dışında çok fazla yöntem geliştirildi ve gerçek hayatta iş süreçlerinde aktif ve verimli bir şekilde kullanıldı ve yenilenen ve gelişen algoritmaları sayesinde de daha efektif ve daha önemli işlerde de kullanılmaya devam etmektedir. Bu güncel algoritmalardan bir tanesi de; 2014'teki tanıtımından bu yana hızla Kaggle'da kullanılan en popüler yöntemlerden biri haline gelen ve 2015'te Kaggle'da yayınlanan 29 meydan okumayı kazanan çözümler arasında tam 17'sinin kullanmış olduğu, bunun dışında 2015 KDD kupasında ise seçilen en iyi 10 çözümün tamamının kullandığı olduğu, "XGBoost" algoritmasıdır (Nielsen, 2016).

XGBoost, tüm sektörlerden araştırmacıların ilgisini çeken, etkili ve önemli bir yöntem olup geçmişi, tecrübesi (yapılan çalışmalar – literatür), içeriği, avantajları ve dezavantajları bilinmesi misyonu, araştırmacılar tarafından önem arz eden yeni (2014), güçlü, etkili ve sağlık sektörü başta olmak üzere birçok karar mekanizması için hayat kurtarıcı olabilecek önemli bir alternatif veri madenciliği yöntemi olarak ön plana çıkmaktadır. Bizlerde bu çalışmada bu misyonu hedeflemiş bulunmaktayız.

2. XGBoost ALGORİTMASI

2.1. Algoritmanın Geçmişi

Aşırı Gradyan Artırma Algoritması, kısa adıyla XGBoost (Extreme Gradient Boosting Algorithm), KA algoritmasının özelleştirilmiş bir formu olup, ilk olarak Tianqi Chen ve Carlos Guestrin tarafından önerilen ve birçok bilim insanının takip eden çalışmalarında sürekli olarak optimize edilen ve geliştirilen bir modeldir. (Li, Yin, Quan & Zhang, 2019; Amjad, Ahmad, Ahmad, Wróblewski, Kamiński, Kamińsk & Amjad, 2022). 2015 yılındaki makine öğrenmesi yarışmalarında ise çok ciddi bir popülerlik kazanmıştır.

2.2. Literatür Taraması

Chen ve Guestrin (2016), XGBoost adını verdikleri ve baştan sona güçlendirme adımlarını içeren eksik veriler için eksikliği duyarlı yeni bir algoritma ve yaklaşık ağaç öğrenimi için ağırlıklı kantil taslağı önerdiklerini ifade etmişlerdir. Ölçeklenebilir bir ağaç güçlendirme sistemi oluşturmak

için önbellek erişim modelleri, veri sıkıştırma ve parçalama hakkında fikirler sunduklarını ve fikirleri birleştirerek, XGBoost modelinin mevcut sistemlerden çok daha az kaynak kullanarak milyarlarca örneğin ötesine ölçeklendiğini belirtmişlerdir.

Ogunleye ve Wang (2020), yüksek performansı korurken daha az özellik kullanan azaltılmış model arzu edildiği için birkaç özellik seçimi yöntemini toplu güçleriyle birleştiren küme teorisine dayalı kural sunduklarını belirtmişlerdir. Uygulamalarını ise dünya nüfusunun %10'unu ve Güney Afrika nüfusunun %15'ini etkileyen bir tehdit olan Kronik Böbrek Hastalığı (KBH), verileri üzerinde gerçekleştirmişler ve bu hastalığın erken ve ucuz, doğru ve güvenilir teşhisinin ise Güney Afrika'da yılda 20.000 hayat kurtarabileceği iddiasında bulunmuşlardır.

Qin, Zhang, Bao, Zhang, Liu ve Liu (2021), parçacık sürüsü optimizasyonuna dayalı bir XGBoost kredi puanlama modeli önermişler ve bu algoritmayı dört farklı veri kümesine uygulamışlar ve sonuçların genel yöntemlerden daha iyi olduğunu raporlamışlardır.

Ma, Zhao, He, Li, Dong, Wang ve Wang (2021), ani sel riskinin değerlendirilmesi için XGBoost modelini tanıtmakta ve ardından optimum etkisini doğrulamak için iki giriş stratejisini ve En Küçük Kareler Destek Vektör Makinesi (Least Square Support Vector Machine - LSSVM) modelini birleştirerek ani sel riski değerlendirmesi için XGBoost tabanlı yöntemi önermişler XGBoost'un %84'le daha iyi sonuç verdiğini, ani sel envanteri tarafından doğrulanan güvenilir ani sel riski haritaları sağladığını yöntemin tespit edilen bir sınırlamasının da "zaman alıcı açgözlü bir algoritma olduğu için çok iş parçacıklı optimizasyonun gerekli olmaması" olduğunu eklemiş ve sonuçların ani sel baskını haritalarının tespitine önemlli bir katkı sağlayabileceği şeklinde olduğunu belirtmişlerdir.

Ramaneswaran, Srinivasan, Vincent ve Hang (2021), mikroskobik beyaz kan hücresi görüntülerinden akut lenfoblastik löseminin (ALL) sınıflandırılması için hibrit bir Inception v3 XGBoost modeli önermişler ve XGBoost modelinin sınıflandırma için iyi bir model olduğunu ve deney sonuçlarının önerilen modelin literatürde tanımlanan diğer yöntemlerden daha iyi performans gösterdiğini belirtmişlerdir. Çalışma Tedavi edilmezse birkaç hafta içinde ölüme neden olabilen, yaşamı tehdit eden bir hastalık Akut Lenfoblastik Lösemi (ALL), tüm pediatrik kanserlerin %25'ini oluşturan en yaygın pediatrik hastaların verileri kullanılarak yürütüldüğünü çalışmalarında ayrıca belirtmişlerdir.

Yotsawat, Wattuya, Srivihok (2021), Bayesçi Hiperparametre Optimizasyonu (XGB-BO) kullanılarak XGBoost sınıflandırıcısına dayalı geliştirilmiş bir kredi puanlama modeli önermişler ve bu modelin üç ayrı veri kümesinde sırasıya %4,10, %3,03 ve %2,76'lık doğruluk iyileştirmesiyle umut verici sonuçlar gösterdiğini ve geleneksel KA, DVM, YSA, Lojistik Regresyon (LR), Rastgele Orman (RO) Ve Torbalama (T) gibi yöntemlerden daha iyi performans gösterdiğini belirtmişlerdir.

Mushava ve Murray (2022), iyi bilinen ve sağlam bir sınıflandırma yöntemi olan XGBoost'ta, genelleştirilmiş aşırı değer dağılımının kantil fonksiyonunun nadir vakaların tespitini geliştirmek için bir bağlantı fonksiyonu olarak kullanılmasını önermişler ve sınıf dengesizliği veri kümesinde kullanmışlar sonuçların iş değeri katacağını ifade etmişlerdir.

Wang, Li, Cheng, Zhou & Li (2022), özellik seçimini LR ile, temerrüt ayrımcılığını ise XGBoost ile gerçekleştirdikleri kişisel kredi riski değerlendirme modeli oluşturmuşlar ve XGBoost'a dayalı kişisel kredi riski değerlendirme modelinin güçlü temerrüt ayrımı yeteneğine ve sağlamlığa sahip olduğu sonucuna ulaşmışlardır.

Zhang, Jia ve Shang (2022), yaptıkları çalışmada XGBoost'un düzenleme terimini optimize etmeye çalışmışlar ve karışık örnekleme ve topluluk öğrenmeye dayalı bir sınıflandırma algoritması önermişlerdir. Temel fikrin, veri işleme aşaması için SVM-SMOTE aşırı örnekleme (over-sampling) ve kolay grup (Easy Ensemble) yetersiz örnekleme (under sampling) teknolojilerini birleştirmek ve ardından eğitim ve topluluk yoluyla XGBoost'a dayalı nihai modeli elde etmek olduğunu ifade etmişlerdir. Bu arada, en önemlisi sınıflandırma tahminini gerçekleştirmek için "en uygun parametrelerin Bayes optimizasyon algoritması aracılığıyla otomatik olarak aranır ve ayarlandığını" belirtmişlerdir. Deneysel aşamada, G-ortalama (G-mean) ve eğri altında kalan alan (Area Under the Curve - AUC) değerleri, kullanılan sınıflandırma modellerinin ve örnekleme yöntemlerinin performansını analiz etmek ve değerlendirmek için kullanmışlar ve deneysel sonuçların, önerilen algoritmanın uygulanabilirliğini ve etkinliğini vurguladığını belirtmişlerdir.

Zhang, Ma, Zhang , Sun, Zhou, Mi ve Wen (2023), SHAP (SHapley Additive explanation)-XGBoost algoritmasına dayalı heyelan duyarlılık değerlendirme modelleri için kapsamlı bir çerçeve oluşturmayı, heyelanı etkileyen faktörlerin bölgesel farklılıklarının ve mekâna bağlı heterojenliği analiz etmeye çalışmışlardır. Heyelanların mekânsal dağılımının heterojen ve karmaşık olduğunu ve heyelanların oluşumunda etkili olan her faktörün katkısı belirgin bölgesel özelliklere ve mekânsal heterojenliğe sahip olduğu

sonucuna varmışlardır. SHAP yöntemini kullanarak XGBoost heyelan duyarlılığı değerlendirme modelinin daha fazla açıklanması, küresel ve yerel değerlendirme birimlerinin bakış açısından, mekânsal heterojenlik nedeniyle çeşitli faktörlerin afetlere ne kadar katkıda bulunduğu farklılıklarının nicel analizine olanak tanıdığını ifade etmişlerdir.

Zedda (2024), LR ve XGBoost yöntemlerinin temerrüt tahmini için etkinliğini test etmiş ve yanlış sınıflandırma için yeni bir endeks geliştirmiştir.

Yuxuan, Shanshan, Lingyi ve Xin (2024), kayıp fonksiyonunu XGBoost modeline dahil ederek yeni bir model geliştirdiklerini ve dört büyük veri kümesi üzerinde yapılan karşılaştırmalı deneyler, önerilen yöntemin mevcut ana akım yöntemlerinden daha üstün olduğunu ve özellikleri etkili bir şekilde çıkarabileceğini ve dengesiz örnekler sorununu çözebileceğini gösterdiğini belirtmişlerdir.

Martinez, Campillo, & Ibañez, (2024), makine öğrenimi algoritmalarının 'kara kutu' doğasıyla sıklıkla ilişkilendirilen etik endişeleri ele aldıkları çalışmalarında; daha yüksek ödeme gücü, karlılık ve azaltılmış borçluluğun daha düşük iş başarısızlığı eğilimi ile ilişkili olduğunu göstermek için ampirik çalışma yapmışlar, sonuçta ise çalışmalarının XGBoost modelinin şeffaflığını ve yorumlanabilirliğini gözsterdiğini belirtmişlerdir.

Özetle XGBoost algoritması, nispeten yeni bir yöntem olmasında karşın literatürde bu konuyla alakalı olarak, hemen her sektörden önemli uygulama örnekleri olan ve eğer gidişata bakılırsa daha popülaritesini bir süre daha koruyacak olan güncel, güçlü, avantajları olan ve dezavantajları optimizasyonla aşılabilen bir sınıflandırma ve tahmin tekniğidir. Sonuçlarının genel itibariyle, geleneksel diğer yöntemlerden daha iyi olduğu birçok çalışmada raporlanmıştır. Diğer algoritmalarla da melez birçok çalışmaya konu olmuş ve bu melez ya da yeni geliştirilmiş modellerin başarımının daha iyi olduğu vurgulanmıştır. Fakat bu melez ya da yeni modellerden hangilerinin literatürde bir araştırma sahasına dönüşeceğini elbette önümüzdeki yıllarda yapılacak olan çalışmalar belirleyecektir.

2.3. Yapısı

XGBoost, denetimli öğrenme (Supervised Learning) gerçekleştiren yapay zekâ yöntemlerinden birisidir. Denetimli öğrenme, yönteme analizden önce doğru olduğu bilinen belli bir miktarda veri (eğitim verisi) yükleyerek, bir bakıma o yöntemi bu doru bilgilerle eğitme sürecidir. Yöntem yeni gözlemleri bu eğitim verisinden edindiği tecrübeden yararlanarak sınıflandırır ya da değerini tahmin eder. XGBoost, daha güçlü modeller üretmek için "güçlendirme" adı verilen bir süreç uygular. Güçlendirmede

yöntem sürekli yeni bir model üretir ve ürettiği her yeni model bir önceki modelin eksikliklerini giderme amacını güder ve sonrasında tüm bu modelleri birleştirir (Mitchell & Frank, 2017).

Yapay zekâ yöntemleri, güçlü matematiksel altyapıları ve çok sayıda yinelenen yani iteratif yöntemlerdir. Yalnız tüm yapay zekâ yöntemlerinin ortak sıkıntısı da işte buradan gelmektedir: aşırı güçten kaynaklı aşırı uyum (over fitting) sorunu. Şöyle ki çok güçlü olan yapay zekâ yöntemleri, iteratif yapıları sayesinde verinin içerisinde yer alan tüm ilişkileri çözmekte, belli bir aşamadan sonra verideki gürültüyü de öğrenmekte hatta veriyi bir bakıma ezberlemektedir. Bu durum eğitim verisine aşırı uyumla sonuçlanır. Bu da modeli eğitim verisine bağımlı hale getirir. Eğitim verisinden kopan, yani yeni bir veri kümesine uygulanan, yöntemin performansı çok düşer. Çünkü yöntemin eğitim kabiliyeti çok yüksek, genelleme kabiliyeti ise buna bağlı olarak düşüktür. Oysa ki iyi bir yöntem bu ikisi arasında ki dengeyi iyi kurabilmelidir.

Aşırı uyum sorunun tersi de yani eksik uyum problemi de (under fitting) söz konusudur. Bu da problemin veri kümesi çok karmaşık fakat kullanılan model çok basit olursa bu model eksik öğrenir ya da öğrenemez ve performansı çok düşük olur. Fakat yapay zekâ yöntemlerinin hemen hemen tamamının en önemli handikabı aşırı uyum sorunudur.

YSA'nda aşırı uyumdan, parametrelerin (gizli katman sayısı, gizli katmandaki nöron sayıları, toplama fonksiyonları, momentum katsayısı vb.) optimizasyonu yolu ile DVM yöntemin de VC katsayısı yardımı ile optimum uzay boyutunu belirlemeye çalışarak, KA yönteminde ise ön budama ya da son budama yöntemleriyle, modelin performansına olumlu etkisi olmayan aşırı dallanmayı budamak suretiyle kaçınılmaktadır. XGBoost aşırı uyumu önlemek için düzenleme (regulation) terimi başta olmak üzere çeşitli yöntemler kullanmaktadır. (Li, Yin, Quan & Zhang, 2019; Amjad, Ahmad, Ahmad, Wróblewski, Kamiński, Kamińsk & Amjad, 2022).

2.4. Parametreleri

YSA'nda olduğu gibi XGboost'ta da en iyi performansı yakalayabilmek için optimize edilmesi gereken bazı parametreler bulunmaktadır. Eğer parametreler iyi optimize edilmezse yöntemin aşırı uyum sergilemesi ihtimali mevcuttur (Liu, Wu, Liu, Li, Hu & Li, 2021). Bu parametreler:

1. *n_tahmin ediciler (n_estimators)*: Eğitim aşamasında yineleme yani iterasyon sayısıdır. Modelin öğrenme yetisini optimize etmek adına önemlidir. Çünkü bu sayının yüksek belirlenmesi aşırı uyuma, düşük belirlenmesi ise eksik uyuma neden olur.

2. *Minimum Çocuk Aralığı (min_child_weight):* Aşırı uyumu önlemek için en küçük yaprak düğümlerinin örnek ağırlıklarının toplamını ifade eder.

3. *Maksimum Derinlik (max_depth):* Adından da anlaşılacağı üzere ağaçları büyütürken izin verilecek olan maksimum derinlik seviyesini ifade eder. Bu sayı ne kadar büyürse yöntemin kompleksitesi o kadar yüksek olacaktır. Yine aşırı uyuma ve eksik uyuma maruz kalmamak için optimizasyonu önemli bir parametredir.

4. *Alt Örnek (subsample):* Eğitim verisinin bir alt kümesiyle modelin eğitilmesini sağlar.

5. *Colsample_Bytree*: Her ağacı oluştururken ki özellik örnekleme oranı.

6. *Öğrenme_Oranı (Learning Rate)*: Adımların ağırlığını ifade eder yani her iterasyonda modelin iyileştirilmesinin ne kadar olacağını belirler.

7. *Alfa (Alpha) (L1 regularization):* L1 düzenleme terimi.

8. *Lamda (Lambda) (L2 regularization):* L2 düzenleme terimi.

9. *Güçlendirici* (*booster):* Öğrenme algoritmasını seçmek için kullanılır. Seçenekler: gbtree, gblinear, dart.

(Li, Yin, Quan & Zhang, 2019).

2.5. Gradyan Artırmadan Farkı

XGBoost'un özünde de GAA yaklaşımı vardır. Ancak, basit GAA ile XGBoost algoritması arasındaki fark, GAA'nda olduğu gibi zayıf öğrenenlerin eklenmesi sürecinin birbiri ardına gerçekleşmemesidir; makinenin CPU çekirdeğinin uygun şekilde kullanılmasıyla daha fazla hız ve performansa yol açan çok iş parçacıklı bir yaklaşım benimsenir. Bunun dışında, eksik veri değerlerinin otomatik olarak işlenmesini, ardından ağaç yapısının paralel hale getirilmesini desteklemek için blok yapısını ve yeni veriler üzerinde önceden oluşturulmuş bir modeli daha da artırabilmek için sürekli eğitim sürecini de içeren seyrek farkındalıklı uygulama vardır. XGBoost'un sınıflandırma ve regresyon ve öngörücü modelleme problemlerinde yapılandırılmış veya tablolu veri kümelerine hâkim olduğu görülmüştür (Ramraj, Uzir, Sunil & Banerjee, 2016). GAA ile Rasgele Orman (RO) arasındaki temel fark ise GAA'nın önceden oluşturulmuş olanları tamamlamak için yeni bir ağaç eklenirken RO'da ağaçların birbirinden bağımsız olarak oluşturulmaktadır (Pan, 2018).

2.6. Avantajları

- Ölçekleme, eksik değerlerin işlenmesi ve normalizasyon vb. özellik mühendisliği gerektirmez sahiptir (Liu, Wu, Liu, Li, Hu & Li, 2021).
- Makinenin CPU çekirdeğinin uygun şekilde kullanılmasıyla daha fazla hız ve performansa yol açan çok iş parçacıklı bir yaklaşım benimsenir (Ramraj, Uzir, Sunil & Banerjee, 2016).
- Sınıflandırma, regresyon ve sıralama problemlerinde kullanılabilir (Ramraj, Uzir, Sunil & Banerjee, 2016).
- Paralel hesaplama sayesinde çok yüksek hıza ve verimliliğe sahiptir (Liu, Wu, Liu, Li, Hu & Li, 2021).

Şeffaf oluşu (Martinez, Campillo, & Ibañez, 2024).

2.7. Dezavantajları

- Karmaşık içeriği,
- Yalnıza sayısal verilere uygulanabilir (Liu, Wu, Liu, Li, Hu & Li, 2021),
- Eğer parametreler optimal olarak ayarlanamazsa aşırı uyum sorunu yaşar (Liu, Wu, Liu, Li, Hu & Li, 2021),

XGboost veri dengesizliği durumunda sınıflandırma etkisi genellikle ideal değildir (Zhang, Jia ve Shang, 2022),

En iyi performansı elde etmek için, modelin dikkatli bir şekilde ayarlanması gerekir (Ogunleye & Wang, 2020).,

XGBoost'un ayarlanması, sahip olduğu hiperparametre sayısı nedeniyle çok zorlu bir görev olabilir (Ogunleye & Wang, 2020).

3. SONUÇLAR VE YORUMLAR

KA yöntemi literatürde altmış yılı aşkın bir süredir bilinen ve sayısız farklı uygulaması yapılan esnek, sağlam, hızlı, güçlü, şeffaf ve aşırı uyumla mücadele edebilen bir yöntem olarak kabul görmüştür. Sonuçlarının çok iyi olması, büyük ve küçük veri kümelerine uygulanabilmesi, varsayım istemediği için pratik ve hızlı bir şekilde uygulanabilmesi yöntemin bilinen yönleri arasındadır.

Diğer tüm yapay zekâ yöntemleri gibi KA tabanlı olarak da çok fazla algoritma geliştirildi. XGboost bu algoritmaların içerisinde en güncel olanıdır ve yarışmalardan açık ara önde çıkarak kendini ispatlamış önemli

bir algoritmadır. Sınıflandırma, regresyon ve sıralama problemlerinde kullanılabilmekte ve CPU üzerinde yük oluşturmamaktadır.

Günümüzde hala, hemen her veri kümesinde tartışmasız en iyi sonuçları veren bir yöntemden bahsetmek mümkün görünmemektedir. Dolayısıyla problemin çözüm arayan gerçek ya da tüzel kişilerin bu algoritmayı da kendi problemine dönük olarak optimize etmek şartıyla kullanarak sonuçlarına erişmeye çalışması yerinde ve faydalı bir yatırım olacaktır. Çünkü analiz sonuçlarındaki %1'lik bir iyileşme bir şirket için milyonlarca dolarlık iyileştirme anlamına gelebileceği gibi, sağlık sektöründeki uygulamalarda yüzlerce insana doğru teşhis konmasına ve tedavisinin yapılmasına olanak tanıyabilir. Deprem, heyelan ve yangın vb. doğal/doğal olmayan afetlerin yeri ve zamanı konusunda tahmin başarısına ve dolayısıyla bu afetlerden daha az mağduriyet yaşanmasını sağlayabilir.

Yöntemin karnesi, literatür tecrübesi, üstünlükleri, ileri kavrayış yaklaşımı, yapılan araştırma neticesinde çok iyi görünmektedir. Yöntemin yeni uygulamalarını takip etmek, sonuçlarda yaşanacak iyileşme noktasında elverişli olacaktır. Bu tarz yöntemlerdeki iyileşmeye dönük yeni algoritma deneyimleri diğer yöntemler için de ilham verici gelişmeler olabilmektedir. KA tabanlı XGboost yönteminin diğer yöntemlere de yol gösterici olması, zaman süreci çerçevesinde, beklenebilir.

Kaynakça

Adeola Ogunleye; Qing-Guo Wang (2020), "XGBoost Model for Chronic Kidney Disease Diagnosis", *IEEE/ACM Transactions On Computational Biology And Bioinformatics,* 17(6), 2131-2140.

Bingyue Pan, (2017), "Application of XGBoost algorithm in hourly PM2.5 concentration prediction", IOP Conference Series: Earth and Environmental Science, Volume 113, 3rd International Conference on Advances in Energy Resources and Environment Engineering 8–10 December 2017, Harbin, China.

Chao Qin, Yunfeng Zhang, Fangxun Bao, Caiming Zhang, Küçük Liu & Peipei Liu, (2021), "XGBoost Optimized by Adaptive Particle Swarm Optimization for Credit Scoring", *Mathematical Problems in Engineering,* 1-18.

Didrik, Nielsen, (2016), "Tree Boosting With XGBoost", Master of Science in Physics and Mathematics, https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2433761/16128_FULLTEXT.pdf

Jialin Liu, Jinfa Wu, Siru Liu, Mengdi Li, Kunçang Hu, Ke Li, (2021). "Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model", Plos One, 16(2), 1-11.

Jiangtao Li, Xingqin An, Qingyong Li, Chao Wang, Haomin Yu, Xinyuan Zhou, Yangli-ao Geng, (2022), "Application of XGBoost algorithm in the optimization of pollutant concentration", Atmospheric Research, 276, 1-16.

Junyi Zhang, Xianglong Ma, Jialan Zhang ,Deliang Sun, Xinzhi Zhou, Changlin Mi, Haijia Wen (2023), "Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model", Journal of Environmental Management, 332, *Journal of Environmental Management,* 332*,* 1-20.

Kui Wang, Meixuan Li, Jingyi Cheng, Xiaomeng Zhou & Gang Li (2022), "Research on personal credit risk evaluation based on XGBoost", *Procedia Computer Science,* 199, 1128-1135.

Maaz Amjad, Irshad Ahmad, Mahmood Ahmad, Piotr Wróblewski, Paweł Kamiński, Paweł Kamińsk &Uzair Amjad (2022), "Prediction of Pile Bearing Capacity Using XGBoost Algorithm: Modeling and Performance Evaluation", *Applied Sciences, 12*(4), 1-24.

Mariano Romero Martinez, José Pozuelo Campillo, & Pedro Carmona Ibañez, (2024), "Ethical transparency in business failure prediction: uncovering the black box of xgboost algorithm", *Spanish Journal Of Finance And Accounting,* https://doi.org/10.1080/02102412.2024.2423498©.

Meihong Ma, Gang Zhao, Bingshun He, Qing Li, Haoyue Dong, Shenggang Wang, Zhonglia Shenggang Wang (2021), "XGBoost-based method

for flash flood risk assessment", Journal of Hydrology 598, https://doi.org/10.1016/j.jhydrol.2021.126382.

Ping Zhang, Yiqiao Jia ve Youlin Shang (2022), "Research and application of XGBoost in imbalanced data", International Journal of Distributed Sensor Networks, 18(6), 1-10.

Ramraj S., Nishant Uzir, Sunil R., & Shatadeep Banerjee, (2016), "International Journal of Control Theory and Applications", International Science Press, 9(40), 651-662.

Rory Mitchell & Eibe Frank (2017), "Accelerating the XGBoost algorithm using GPU computing", *PeerJ Computer Science*, 1-37.

S. Ramaneswaran, Kathiravan Srinivasan, Başbakan Durai Raj Vincent, Chuan-Yu Chang (2021), "Hybrid Inception v3 XGBoost Model for Acute Lymphoblastic Leukemia Classification", Computational and Mathematical Methods in Medicine, 1-10. https://onlinelibrary.wiley.com/doi/epdf/10.1155/2021/2577375.

Stefano Zedda (2024), "Credit scoring: Does XGBoost outperform logistic regression? A test on Italian SMEs", *Research in International Business and Finance,* 70, 1-28.

Tianqi Chen , Carlos Guestrin (2016), "XGBoost: A Scalable Tree Boosting System", KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785 – 794, https://doi.org/10.1145/2939672.29397.

Wei Li, Yanbin Yin, Xiongwen Quan & Han Zhang (2019), "Gene Expression Value Prediction Based on XGBoost Algorithm", Sec. Computational Genomics, 10, https://doi.org/10.3389/fgene.2019.01077.

Wirot Yotsawat, Pakaket Wattuya, Anongnart Srivihok (2021), "Improved credit scoring model using XGBoost with Bayesian hyper-parameter optimization", *International Journal of Electrical and Computer Engineering (IJECE),* 11(6), 5477-5487.

Xia, Yuxuan; Jiang, Shanshan; Meng, Lingyi & Ju, Xin (2024), "XGBoost-B-GHM: An Ensemble Model with Feature Selection and GHM Loss Function Optimization for Credit Scoring", *Systems*,12(7), 254-264.

Vasif Nabiyev, (2021), "Yapay Zeka", Seçkin Yayıncılık: İstanbul.

Yunus Mushava &Michael Murray (2022), "A novel XGBoost extension for credit scoring class-imbalanced data combining a generalized extreme value link and a modified focal loss function", *Expert Systems with Applications,* 202, 1-17.