

Yönetim Bilişim Sistemleri Perspektifinden Algoritmik Yanlılık ve Etik Karar Verme

Üzeyir Fidan¹

Özet

Yapay zekâ (YZ) ve veri odaklı karar destek sistemleri, kurumsal ve toplumsal düzeyde karar alma süreçlerini dönüştürerek daha verimli ve hızlı sonuçlar elde edilmesini sağlarken, algoritmik yanlılık önemli bir etik ve yönetim sorunu olarak öne çıkmaktadır. Algoritmaların eğitim verilerinde bulunan dengesizlikler, modelleme süreçlerinde yapılan tercihler ve KDS'deki şeffaflık eksikliği, belirli gruplara karşı sistematik olarak adaletsiz sonuçlar doğurabilmektedir. Finans, insan kaynakları, sağlık ve hukuk gibi kritik alanlarda, algoritmaların veri temelli karar süreçlerine entegrasyonu, veriye dayalı objektif kararlar üretme hedefi taşırken, aksine önyargılı modelleme süreçleri nedeniyle mevcut eşitsizlikleri derinleştirme riski taşımaktadır. Bu çalışmanın amacı, algoritmik yanlılığın veri yanlılığı, model yanlılığı ve karar destek sistemlerinde yanlılık başlıkları altında incelenmesi, karar verme teorileri (sınırlı rasyonellik, beklenti teorisi ve DSS teorisi) çerçevesinde değerlendirilmesi ve etik, hesap verebilir ve şeffaf karar destek mekanizmalarının oluşturulması için stratejik öneriler sunulmasıdır. Yönetim Bilişim Sistemleri (YBS) perspektifinden bakıldığında, algoritmik yanlılık yalnızca teknik bir sorun değil, aynı zamanda organizasyonların etik sorumluluğuyla doğrudan ilişkili bir yönetim problemidir. Bu bağlamda FAT (Fairness, Accountability, Transparency) çerçevesi, algoritmaların adil, hesap verebilir ve şeffaf hale getirilmesini sağlayarak karar mekanizmalarının güvenilirliğini artırmada kritik bir rol oynamaktadır. Kurumsal düzeyde algoritmik denetim mekanizmalarının geliştirilmesi, açıklanabilir yapay zekâ (AYZ) tekniklerinin yaygınlaştırılması ve etik regülasyonlara dayalı yönetim modellerinin benimsenmesi, algoritmik yanlılığın yönetimi ve adil yapay zekâ uygulamalarının geliştirilmesi açısından temel çözüm alanlarıdır. Sonuç olarak, algoritmik sistemlerin yalnızca teknik optimizasyonlarla değil, aynı zamanda etik ve yönetsel çerçevelerle desteklenmesi, YBS kapsamında sorumlu yapay zekâ kullanımının yaygınlaştırılması için kaçınılmaz bir gereklilik olarak değerlendirilmektedir.

1 Dr., Uşak Üniversitesi, uzeyir.fidan@usak.edu.tr, <https://orcid.org/0000-0003-3451-4344>

1. Giriş

Son yıllarda YZ ve veri odaklı sistemlerin hızla yaygınlaşması, karar verme süreçlerinde algoritmaların kritik bir rol oynamasına neden olmuştur. Algoritmalar, büyük ölçekli verileri analiz ederek bireyler, kurumlar ve toplumlar için stratejik kararları destekleyen sistemlere dönüşmüştür (Çam, 2024; Lepri vd., 2018). Ancak, bu süreçte algoritmik yanlılık önemli bir sorun olarak ortaya çıkmaktadır.

Algoritmik yanlılık, bir algoritmanın belirli gruplara veya bireylere sistematik olarak önyargılı veya adaletsiz sonuçlar üretmesi durumudur (Barocas vd., 2019). Bu yanlılık, çoğunlukla eğitim verilerindeki dengesizlikler, model tasarımındaki hatalar ve algoritmik karar süreçlerinde kullanılan metodolojilerden kaynaklanmaktadır. Özellikle sağlık, finans, insan kaynakları ve hukuk gibi kritik alanlarda, algoritmik yanlılık sosyal eşitsizlikleri artırarak toplumsal adalet anlayışını zedelemektedir (O'Neil, 2016).

YBS perspektifinden bakıldığında, algoritmik yanlılık, iş zekâsı (İZ), karar destek KDS ve veri analitiği (VA) süreçlerinde doğrudan etkili olup, kurumsal karar verme mekanizmalarının güvenilirliği ve etik boyutu açısından ele alınması gereken temel bir sorundur (Shapiro ve Varian, 1999). Bu bağlamda, algoritmik yanlılığın tanımlanması, kaynaklarının incelenmesi ve etik sorumluluk çerçevesinde çözümler üretilmesi büyük önem taşımaktadır.

YBS, organizasyonların bilgi tabanlı karar alma süreçlerini optimize eden sistemler bütünüdür (Erbey 2024; Laudon ve Laudon, 2017). Bu sistemler, KDS ve İZ araçlarını kullanarak yöneticilere stratejik içgörüler sağlamaktadır. Ancak, bu sistemlerin kullandığı algoritmalar belirli önyargıları içerebilmekte ve bu da karar kalitesini doğrudan etkilemektedir.

Özellikle aşağıdaki faktörler, YBS içinde algoritmik yanlılık riskini artırmaktadır:

- **Veri Kaynaklarının Heterojenliği:** Çeşitli kaynaklardan gelen veriler, temsil sorunlarına ve eğilim bozukluklarına yol açabilmektedir (Mehrabi vd., 2021).
- **Modelleme Süreçlerinde Kullanılan Algoritmalar:** Kullanılan makine öğrenimi modelleri belirli gruplara karşı ayrımcılık yapabilmektedir (Zliobaitė, 2017).
- **KDS'nin Şeffaflık Eksikliği:** Algoritmaların “kara kutu” niteliği, yöneticilerin sonuçları anlamasını zorlaştırabilmektedir (Lipton, 2018).

Bu bağlamda, KDS ve İZ uygulamalarının etik boyutunu ele almak, organizasyonların sorumlu YZ politikalarını benimsemeleri açısından kritik bir gereklilik haline gelmektedir (Danks ve London, 2017; Tuna ve Görmez, 2024).

Algoritmik yanlılık, yalnızca teknik bir sorun değil, aynı zamanda etik ve toplumsal sonuçlar doğuran bir olgudur (Binns, 2018). Bu nedenle, YBS kapsamında karar verme süreçlerinin adalet, hesap verebilirlik ve şeffaflık (FAT - Fairness, Accountability, Transparency) çerçevesinde ele alınması büyük önem taşımaktadır (Mittelstadt vd., 2016).

Bu bağlamda, algoritmaların farklı kurumsal ve toplumsal alanlarda yarattığı etkiler önemli tartışmalara yol açmaktadır. Örneğin, insan kaynakları yönetiminde, işe alım süreçlerinde kullanılan algoritmaların cinsiyet, etnik köken veya sosyo-ekonomik statüye dayalı ayrımcılık yapması olası bir risk olarak karşımıza çıkmaktadır (Raghavan vd., 2020). Benzer şekilde, finans sektöründe kredi tahsis süreçlerinde algoritmik ayrımcılığın düşük gelirliler veya belirli etnik grupları olumsuz etkilemesi, finansal eşitsizlikleri artırabilir (Ustun ve Rudin, 2019). Pazarlama ve müşteri analitiği alanında ise kişiselleştirilmiş reklam algoritmalarının belirli grupları yanlış veya ayrımcı kategorilere yerleştirmesi, tüketici hakları açısından sorun teşkil edebilmektedir (Tschantz vd., 2015).

Öte yandan, algoritmaların toplumsal sistemlerdeki kullanımı da belirli gruplar üzerinde ayrımcı sonuçlar doğurabilmektedir. Örneğin, yargı sistemlerinde kullanılan yapay zeka destekli karar mekanizmalarının, belirli grupları daha yüksek risk kategorisine sokarak adaletin tarafsızlığını zedeleme riski bulunmaktadır (Angwin vd., 2022). Benzer şekilde, tıbbi teşhis ve tedavi algoritmalarının belirli etnik veya sosyo-ekonomik grupları yeterince temsil etmemesi, sağlık hizmetlerine erişimde eşitsizlikleri artırabilir (Obermeyer vd., 2019). Ayrıca, sosyal medya ve haber platformlarında yanlış içeriklerin öne çıkarılması, bilgi manipülasyonu riskini güçlendirmektedir (Bakshy vd., 2015). Eğitim alanında ise kişiselleştirilmiş içerikler ve çevrimiçi sınav değerlendirme araçlarının, öğrenci performansını geçmiş verilere veya kısmi demografik bilgilere dayanarak tahmin etmesi, yeterince temsil edilmeyen sosyo-ekonomik veya kültürel grupları dezavantajlı konuma düşürebilmektedir (Yılmaz, 2022).

Tüm bu örnekler, algoritmik yanlılığın sadece teknik bir sorun olmadığını, aksine organizasyonların etik sorumluluğunun bir parçası olarak ele alınması gerektiğini göstermektedir. Bu kitap bölümü, algoritmik yanlılığın YBS perspektifinde ele alınarak veri yanlılığı, model yanlılığı ve karar süreçleri

bağlamında incelenmesini, ayrıca karar verme teorileri çerçevesinde değerlendirilmesini amaçlamaktadır.

2. Algoritmik Yanlılık: Kavramsal Çerçeve ve Kategoriler

Algoritmalar, büyük veri analitiği, makine öğrenimi ve YZ tabanlı sistemler yoluyla insan kararlarını desteklemek için yaygın bir şekilde kullanılmaktadır. Ancak, bu sistemlerin verdiği kararların adil, tarafsız ve etik olup olmadığı kritik bir sorundur (Mehrabi vd., 2021). Algoritmik yanlılık, sistematik bir şekilde belirli gruplara veya bireylere karşı ayrımcılık yapılmasına neden olan hatalı veri kullanımı, modelleme süreçleri veya karar mekanizmalarından kaynaklanabilmektedir (Barocas vd., 2023).

Algoritmik yanlılık YBS bağlamında üç temel kategoride incelenmektedir:

- Veri Yanlılığı (Data Bias): Kullanılan veri setlerinin eğilimli olması nedeniyle ortaya çıkar.
- Model Yanlılığı (Model Bias): Algoritmanın tasarımından ve modelleme sürecinden kaynaklanan hatalardır.
- KDS'de Algoritmik Yanlılık (DSS Bias): Algoritmaların YBS içindeki karar destek süreçlerini nasıl etkilediğiyle ilgilidir.

2.1. Veri Yanlılığı

Veri, YZ ve makine öğrenimi modellerinin temel yapı taşıdır. Ancak, veri toplama, işleme ve modelleme aşamalarında çeşitli yanlılık türleri oluşabilmektedir (Žliobaitė, 2017). Veri yanlılığı, algoritmik sistemlerin belirli grupları sistematik olarak yanlış sınıflandırmasına veya marjinalize etmesine neden olabilmektedir.

2.1.1. Veri Kaynaklarında Yanlılık

Kullanılan veri setleri, toplumun belirli kesimlerini temsil etme hususunda zayıf kalabilmekte hatta hiç temsil etmeyebilmektedir. Örneğin, yüz tanıma sistemlerinde genellikle beyaz ten rengine sahip bireylerin daha fazla temsil edilmesi, sistemin diğer etnik gruplar üzerinde daha az doğru çalışmasına neden olmaktadır (Buolamwini ve Gebru, 2018).

Veri kaynaklarında yanlılığın oluşmasına neden olan diğer bir husus ise tarihi verilerden kaynaklanmaktadır. Eğer bir kurum geçmişte ayrımcı bir işe alım politikası izlemişse, bu politika geçmiş veriye yansımakta ve algoritma, benzer ayrımcı kararları tekrarlayabilmektedir (Barocas ve Selbst, 2016).

2.1.2. Örneklem ve Temsiliyet Sorunları

Veri kaynaklarında olduğu gibi elde edilen örnekleme toplumun belirli kesimlerini temsil noktasında yetersiz olabilmektedir. Örneğin, kadınların teknoloji sektöründeki temsil oranı düşükse, algoritmalar kadın adayları daha düşük olasılıkla işe alabilmektedir (Caliskan vd., 2017). Öte yandan belirli bir grup veya olayın veri setinde fazla temsil edilmesi, algoritmaların bu durumu genellemesine neden olabilmektedir.

2.1.3. Veri Ön İşleme Süreçlerinde Karşılaşılan Yanlılıklar

Gerçek hayat verileri genellikle eksik veya tutarsızdır. Bu aşamada sıklıkla veri ön işleme süreçleri ile veri kümesi temizlenmekte veya eksik veriler tamamlanmaktadır (Gündüz ve Polat, 2021). Bu yöntemler her ne kadar veriyi işleme sürecine katkı sağlasa da yanlılık oluşmasına da zemin hazırlamaktadır. Örneğin, eksik değerleri varsayılan bir ortalama ile doldurmak, belirli grupların davranışlarını değiştirebilmektedir (Mehrabi vd., 2021).

Modelin eğitilmesinde kullanılan değişkenler, belirli bir grubu dezavantajlı hale getirebilmektedir. Örneğin, geçmişte kredi başvurularını reddeden bir sistemin eğitiminde kullanılan gelir seviyesi değişkeni, belirli demografik grupları sistematik olarak dışlayabilmektedir (Ustun ve Rudin, 2019).

2.2. Model Yanlılığı

Algoritmalar, model tasarımı, öğrenme süreci ve optimizasyon metodolojileri nedeniyle yanlılık gösterebilmektedir. Bu yanlılık türü genellikle kullanılan makine öğrenimi teknikleri ve algoritmaların eğitimi sırasında ortaya çıkmaktadır (Kim vd., 2019).

2.2.1. Algoritmik Modelleme ve Önyargılar

Algoritmik modelleme sürecinde, algoritmanın eğitimi ve adalet anlayışıyla ilgili iki temel sorun öne çıkmaktadır. Bu süreçte kullanılan verilerin niteliği ve modelin karar verme mekanizmaları hem performans açısından hem de etik açıdan önemli sonuçlar doğurmaktadır.

2.2.1.1. Önyargılı Model Eğitimi

Modelin sadece belirli bir demografik gruba dayalı olarak eğitilmesi, karar süreçlerini bu grubun özelliklerine göre şekillendirebilmektedir (Zafar vd., 2017).

2.2.1.2. Doğruluk-Adalet Dengesi

Algoritmalar çoğu zaman doğruluk oranını artırmak için belirli grupları feda edebilmekte, bu da adalet sorunlarına yol açmaktadır (Dwork vd., 2012).

2.2.2. Model Seçimi ve Optimizasyon Sürecinde Yanlılık

Model seçimi ve optimizasyon sürecinde, kayıp fonksiyonu seçimi ve aşırı öğrenme (overfitting) olmak üzere iki temel sorun öne çıkmaktadır. Bu sorunlar, modelin genelleme yeteneğini ve karar verme doğruluğunu doğrudan etkileyerek algoritmanın performansını belirleyici unsurlar arasında yer almaktadır.

2.2.2.1. Kayıp Fonksiyonu Seçimi

Kayıp fonksiyonları genellikle modelin ortalama hatasını minimize etmeye çalışır. Ancak, belirli grupların hata oranı diğerlerinden yüksek olabilmektedir (Menon ve Williamson, 2018).

2.2.2.2. Aşırı Öğrenme

Modelin, eğitim verisini aşırı derecede öğrenmesi, genelleme kabiliyetini düşürerek, yeni veriler üzerinde yanlış tahminler yapmasına neden olabilmektedir.

2.2.3. Modelin Gerçek Dünya Verileri ile Uyumluluğu

Modelin gerçek dünya verileri ile uyumluluğu sürecinde, veri dağılımının değişmesi ve gerçek dünya testleri olmak üzere iki temel sorun öne çıkmaktadır. Bu sorunlar, modelin eğitim sırasında öğrendiği örüntülerin gerçek dünya senaryolarında ne ölçüde geçerli olduğunu belirleyerek genel performansını ve güvenilirliğini doğrudan etkilemektedir.

2.2.3.1. Veri Dağılımının Değişmesi

Model eğitildiği veriyle uyumlu olabilir, ancak gerçek dünya verileri değiştikçe model performansı düşebilir ve yanlılık artabilmektedir (Mehrabi vd., 2021).

2.2.3.2. Gerçek Dünya Testleri

Modelin farklı demografik gruplar üzerindeki performansı test edilmediğinde, yanlış tahminler yapması kaçınılmazdır (Obermeyer vd., 2019).

2.3. Karar Destek Sistemlerinde Algoritmik Yanlılık

KDS, büyük veri analitiği ve YZ destekli modeller kullanarak yöneticilere veri odaklı kararlar alma konusunda rehberlik etmektedir. Ancak, bu sistemlerin algoritmik süreçlere dayalı olması, belirli önyargıların karar alma mekanizmalarına yansımaya neden olabilmektedir. Veri setlerindeki dengesizlikler, modelleme süreçlerindeki eksiklikler ve karar mekanizmalarının şeffaf olmaması, algoritmaların belirli gruplara karşı adaletsiz kararlar vermesine yol açmaktadır. Bu bağlamda, öncelikle KDS'nin algoritmalarla nasıl bütünleştiği ve kurumsal yönetim süreçlerinde nasıl bir rol oynadığının incelenmesi önem arz etmektedir.

2.3.1. Yönetim Süreçlerinde Karar Destek Sistemleri ve Algoritmaların Rolü

KDS, organizasyonların veri odaklı ve analitik tabanlı kararlar almasını sağlayan bilişim sistemleridir (Laudon ve Laudon, 2017). YBS kapsamında ele alındığında, KDS, İZ, büyük veri analitiği ve otomasyon tabanlı karar mekanizmalarını kullanarak yöneticilere stratejik karar alma süreçlerinde destek sunmaktadır (Shapiro ve Varian, 1999). Özellikle finans, insan kaynakları, sağlık ve tedarik zinciri yönetimi gibi alanlarda, algoritmalar yöneticilere veri analizleri aracılığıyla tahminlerde bulunmakta, alternatif senaryolar oluşturmada ve optimum kararları önermektedir (Power vd., 2015).

Ancak, bu süreçte kullanılan algoritmalar belirli gruplara karşı yanlılık içerebilmektedir. Örneğin, iş gücü planlaması için kullanılan bir KDS, geçmiş verilere dayalı olarak cinsiyet veya etnik köken farklılıklarını dikkate almadan tahminlerde bulunabilmektedir (Raghavan vd., 2020). Eğer bu sistem, geçmişte kadınların teknik pozisyonlarda daha az yer aldığı bir veri setiyle eğitildiyse, gelecekteki işe alım süreçlerinde de benzer eğilimleri sürdürmektedir. Bu tür önyargılar, sadece iş gücü yönetiminde değil, aynı zamanda finansal kararlar, müşteri risk değerlendirmeleri ve kredi skorlama sistemleri gibi birçok farklı alanı doğrudan etkilemektedir (Barocas vd., 2023).

Bu nedenle, yönetim süreçlerinde başarılı KDS'nin geliştirilebilmesi, algoritmaların nasıl eğitildiği, hangi verilerin kullanıldığı ve bu verilerin ne kadar temsil edici olduğu gibi faktörlere bağlıdır (Mehrabi vd., 2021). Şeffaf olmayan veya “kara kutu” olarak adlandırılan modellerin kullanımı, yöneticilerin algoritmik kararları nasıl değerlendireceğini anlamasını zorlaştırmakta ve sistematik ayrımcılık riskini artırmaktadır (Lipton, 2018).

2.3.2. Algoritmik Yanlılığın Kurumsal Karar Verme Süreçlerindeki Etkileri

Algoritmik yanlılık, kurumsal karar alma mekanizmalarında büyük etkiler yaratmaktadır. Özellikle finansal sistemlerde kredi skorlama algoritmalarında görülen yanlılıklar, belirli sosyo-ekonomik grupların krediye erişimini zorlaştırabilmektedir (Ustun ve Rudin, 2019). Geleneksel olarak düşük gelirli bireyler veya belirli bölgelerde yaşayan gruplar, kredi onay süreçlerinde daha düşük olasılıkla kabul edilmektedir. Eğer bir banka, geçmiş verilerine dayanarak kredi değerlendirme algoritmasını oluşturursa ve bu veriler halihazırda belirli gruplara karşı ayrımcılık içeren kredi verme politikalarına sahipse, sistem mevcut eşitsizlikleri pekiştirerek dezavantajlı bireylerin daha da zorlanmasına neden olabilmektedir (O’Neil, 2016). Benzer bir durum insan kaynakları yönetimi alanında da gözlemlenmektedir. Büyük şirketler, iş başvurularını değerlendirmek için YZ destekli algoritmaları kullanmaktadır. Ancak, geçmiş işe alım verileri cinsiyet veya etnik köken açısından önyargılıysa, bu durum gelecekteki işe alım süreçlerinde de kendini göstermektedir (Caliskan vd., 2017). Örneğin, teknik pozisyonlara başvuran kadın adaylar geçmişte daha az işe alınmışsa, bir makine öğrenimi modeli bunu bir norm olarak öğrenmekte ve erkek adayları daha fazla tercih eden kararlar vermektedir. Amazon’un geçmişte işe alım sürecinde cinsiyet önyargılı bir algoritma kullandığı ve bu algoritmanın erkek adayları daha fazla öne çıkardığı rapor edilmiştir (Dastin, 2018). Bu tür hatalar, insan kaynakları yönetimi süreçlerinde fark edilmezse, kurumsal çeşitlilik ve kapsayıcılık politikaları ile çelişen sonuçlar ortaya çıkmaktadır.

Bunun yanı sıra, müşteri analitiği ve pazarlama alanlarında da algoritmik yanlılık önemli sonuçlar doğurmaktadır. Kişiselleştirilmiş reklam algoritmaları, belirli demografik grupların geçmiş çevrimiçi davranışlarını analiz ederek kullanıcıya özel içerikler sunmaktadır. Ancak, bu sistemler bazen belirli cinsiyet veya etnik kökene sahip kullanıcıları yanlış kategorilere yerleştirebilmektedir (Tschantz vd., 2015). Örneğin, siyahi kullanıcıların daha düşük kredi limitli finansal ürünlerle hedeflenmesi, algoritmanın geçmiş verilerden öğrendiği ayrımcı bir modelin sonucu olarak ortaya çıkmaktadır. Bu gibi önyargılar hem müşteri deneyimini olumsuz etkileyebilmekte hem de şirketlerin marka itibarına zarar verebilmektedir.

2.3.3. Algoritmik Yanlılığı Azaltmaya Yönelik Karar Destek Mekanizmaları

Algoritmik yanlılığın etkilerini azaltmak için çeşitli teknik ve stratejiler geliştirilmektedir. Bunlardan biri, adil modelleme teknikleri olarak bilinen

yaklaşımlardır. Örneğin, “adversarial debiasing” olarak adlandırılan yöntem, bir modelin belirli bir değişken (örneğin, cinsiyet veya etnik köken) ile ilişkilendirilmemesi için öğrenme sürecine eklenen optimizasyon yöntemlerini içermektedir (Zafar vd., 2017). Bu sayede, modelin çıktılarının belirli gruplar lehine veya aleyhine sistematik olarak önyargılı olmaması sağlanmaktadır. Bunun yanında, YBS’de algoritmik denetim ve şeffaflık mekanizmalarının uygulanması, yöneticilerin bu sistemlerin nasıl çalıştığını anlamalarına yardımcı olabileceği düşünülmektedir. AYZ teknikleri, algoritmaların nasıl kararlar aldığına dair şeffaflık sağlayarak, yöneticilere sistematik önyargıları tespit etme ve müdahale etme fırsatı sunmaktadır (Lipton, 2018).

Son olarak, yasal düzenlemeler ve etik standartlar, algoritmik sistemlerin toplum üzerindeki olumsuz etkilerini minimize etmek için önemli bir araçtır. Avrupa Birliği’nin Genel Veri Koruma Yönetmeliği (GDPR) ve ABD’de geliştirilen YZ Etik Rehberleri, algoritmik karar süreçlerinde hesap verebilirliği artırmayı amaçlayan önemli regülasyonlar arasındadır (Mittelstadt vd., 2016).

3. Karar Verme Süreçlerinde Algoritmik Yanlılık ve Teorik Temeller

Karar verme süreçleri, bireylerin veya kurumların mevcut bilgiyi değerlendirerek en uygun seçeneği belirleme sürecini ifade etmektedir. Geleneksel karar teorileri, bireylerin tamamen rasyonel davrandığını varsaymakta ve optimal sonuçlara ulaşmak için gerekli tüm bilgilere sahip olduklarını öngörmektedir. Ancak Herbert Simon’un (1955) geliştirdiği Sınırlı Rasyonellik Teorisi, karar alıcıların rasyonelliklerinin bilgi eksikliği, bilişsel sınırlamalar ve zaman kısıtlamaları nedeniyle sınırlı olduğunu öne sürmektedir (Augier, 2001).

Algoritmalar, KDS’de insan kararlarını destekleyerek daha veri odaklı ve sistematik kararlar alınmasını sağlamaktadır. Ancak, algoritmalar da tıpkı insanlar gibi belirli sınırlamalara sahiptir. Algoritmik kararlar, kullanılan veri, modelleme süreci ve belirlenen hedefler çerçevesinde şekillendiğinden, algoritmaların sınırlı rasyonellik içinde nasıl konumlandığı ve kurumsal karar süreçlerinde nasıl optimize edilebileceği kritik bir araştırma alanı haline gelmiştir.

3.1. Sınırlı Rasyonellik Teorisi ve Algoritmik Kararlar

3.1.1. İnsan Karar Verme Süreçlerinde Sınırlı Rasyonellik

Klasik ekonomi teorileri, bireylerin tamamen rasyonel olduğunu ve kararlarını optimize etmek için tüm alternatifleri değerlendirdiğini

varsaymaktadır (Von Neumann ve Morgenstern, 1944). Ancak Simon, gerçek dünyada karar alıcıların sınırlı bilgi, zaman baskısı ve bilişsel kapasite eksikliği nedeniyle en iyi kararı vermek yerine “yeterince iyi” (satisficing) kararlara yöneldiğini belirtmiştir. Örneğin, bir yönetici yatırım yaparken tüm finansal verileri analiz etme kapasitesine sahip olmayabilmektedir. Bunun yerine, geçmiş deneyimlerine ve mevcut piyasa eğilimlerine dayalı olarak en iyi görünen seçeneği tercih etme eğilimindedir. Aynı durum, algoritmaların karar verme süreçleri için de geçerlidir. Bir algoritma, karar vermek için geniş bir veri kümesine erişebilir ancak, verilerin eksik veya önyargılı olması, karar süreçlerini etkileyerek sınırlı rasyonellik sergileyen sonuçlar üretebilmektedir (Gigerenzer ve Selten, 2001).

3.1.2. Algoritmaların Sınırlı Rasyonellik İçinde Konumlandırılması

Algoritmalar, geniş veri setlerini analiz edebilmekte, kalıpları tanıyabilmekte ve karar süreçlerini hızlandırabilmektedir. Ancak, algoritmaların da veri kısıtlamaları, modelleme hataları ve programlanmış optimizasyon hedefleri nedeniyle sınırlı rasyonellik içinde çalıştığı söylenebilir (Rahwan vd., 2019). Örneğin, bir sigorta şirketi, müşteri primlerini belirlemek için bir algoritma kullanıyorsa, sistem geçmiş müşteri verilerini temel alarak tahminler yapacaktır. Ancak, algoritmanın kullandığı veri seti sadece belirli müşteri gruplarını içeriyorsa veya geçmişte ayrımcı sigorta politikaları izlenmişse, algoritmanın çıktıları da bu doğrultuda yanlı ve sınırlı rasyonellik içerecektir (Barocas vd., 2019).

Algoritmaların karar verme süreçlerinde sezgisel veya yaratıcı çözümler üretememesi, onların sınırlı rasyonellik çerçevesinde değerlendirilmesine yol açmaktadır. İnsan karar vericiler, sınırlı bilgiye sahip olsalar bile deneyimlerinden ve bağlamsal bilgiden yararlanarak yaratıcı çözümler üretebilirler. Oysa ki, algoritmalar genellikle belirli kurallar, optimizasyon fonksiyonları ve veri sınırlamaları dahilinde karar vermektedir (Simon, 1997). Bununla birlikte, algoritmaların karar verme süreçlerinde kendi öğrenme süreçlerinden kaynaklanan yanlılıkları yeniden üretebileceği ve bazı durumlarda hataları fark edemeyeceği de unutulmamalıdır (Žliobaitė, 2017). Bu nedenle, algoritmaların eğitildiği veri setleri ve optimizasyon süreçleri titizlikle incelenmeli, belirli etik kurallar ve denetim mekanizmaları uygulanmalıdır.

3.1.3. Kurumsal ve Stratejik Karar Verme Süreçlerinde Algoritmik Rasyonalite

Algoritmaların sınırlı rasyonellikle çalışması, yönetim süreçlerinde karar destek mekanizmalarının nasıl tasarlanması gerektiğine dair kritik

sorular ortaya çıkarmaktadır. Kurumsal ve stratejik karar alma süreçlerinde, algoritmaların sağladığı rasyonalite, insan faktörüyle bütünleşik bir şekilde ele alınmalı ve bu denge, kararların etkinliği ve adaleti açısından titizlikle değerlendirilmelidir.

Kurumsal düzeyde, algoritmalar özellikle risk yönetimi, finansal tahminleme, müşteri segmentasyonu ve tedarik zinciri optimizasyonu gibi alanlarda kullanılmaktadır (Davenport ve Harris, 2017). Ancak, bu süreçlerde kullanılan YZ sistemleri sadece matematiksel optimizasyon modellerine dayanırsa, işletmelerin uzun vadeli stratejik hedeflerine uyum sağlayamayan kararlar üretebilmektedir (Brynjolfsson ve McAfee, 2017). Örneğin, bir şirketin insan kaynakları departmanı, çalışanların terfi sürecini belirlemek için bir makine öğrenimi modeli kullanabilir. Ancak, bu model yalnızca geçmiş performans verilerini ve iş başarı metriklerini dikkate alarak yönetim yetenekleri, takım çalışması gibi daha soyut unsurları göz ardı edebilmektedir. Sonuç olarak, şirketin stratejik hedefleri doğrultusunda insan sermayesini geliştirmeye yönelik uzun vadeli bir yaklaşım benimsemesi engellenmiş olacaktır (Raisch ve Krakowski, 2021). Benzer şekilde, finans sektöründe kullanılan algoritmik ticaret sistemleri, yüksek frekansta işlem yaparak kısa vadeli kârı en üst düzeye çıkarabilmekte, ancak uzun vadeli ekonomik sürdürülebilirlik açısından ciddi riskler barındırmaktadır (Chorafas, 2011). Bu durum, piyasa istikrarı ve regülasyon politikaları açısından önemli tartışmaları da beraberinde getirmektedir. Bu tür sistemlerin sadece anlık kazançları optimize etmesi, işletmelerin etik, sosyal sorumluluk ve uzun vadeli stratejilerle ilgili hedeflerini ikinci plana atmasına neden olmaktadır. Bu durum, kurumsal KDS'de algoritmaların rasyonel, ancak bağlamsal ve etik değerlere duyarlı olacak şekilde tasarlanması gerektiğini göstermektedir. AYZ modelleri ve insan-merkezli KDS, bu tür riskleri azaltmak için geliştirilen yaklaşımlar arasında yer almaktadır (Doshi-Velez ve Kim, 2017).

3.2. Beklenti Teorisi ve Algoritmik Risk Değerlendirmesi

Karar verme süreçlerinde risk algısı, bireylerin ve organizasyonların nasıl seçim yaptığı üzerinde önemli bir rol oynamaktadır. Daniel Kahneman ve Amos Tversky (1979) tarafından geliştirilen Beklenti Teorisi, bireylerin kararlarını mutlak fayda yerine kazanım ve kayıp perspektifinden değerlendirdiğini öne sürmektedir. Bu teoriye göre insanlar, kazançlar ve kayıpları farklı şekillerde değerlendirmekte ve genellikle kayba karşı aşırı duyarlı davranmaktadırlar (Kahneman ve Tversky, 1979).

Geleneksel rasyonel seçim teorileri, bireylerin optimal kararları vermek için tüm olasılıkları ve riskleri objektif bir şekilde değerlendirdiğini

varsaymaktadır. Ancak Beklenti Teorisi, karar vericilerin duygusal, bilişsel ve psikolojik faktörlerden etkilendiğini ve risk tercihlerini buna göre şekillendirdiğini göstermektedir. Bu çerçevede, algoritmaların risk yönetiminde nasıl bir rol oynadığı ve algoritmik yanlılığın risk değerlendirme süreçlerine nasıl yansıdığı önem kazanmaktadır. Algoritmalar, risk değerlendirme süreçlerinde büyük ölçekli verileri analiz ederek rasyonel tahminler yapabilmekte; ancak algoritmik modelleme süreçleri, eğitim verilerindeki yanlılıklar ve optimizasyon hedefleri nedeniyle risk algısında sistematik önyargılar yaratabilmektedir (Barocas vd., 2019).

3.2.1. Risk Algısı ve Karar Mekanizmaları

Risk algısı, bireylerin ve organizasyonların belirsizlik içeren durumlarla nasıl başa çıktığını belirleyen kritik bir faktördür. Geleneksel modeller, bireylerin riskleri olasılık dağılımları üzerinden rasyonel bir şekilde değerlendirdiğini öne sürerken, Beklenti Teorisi insanların kazançlardan çok kayıplara odaklandığını ve riskten kaçınma veya aşırı risk alma davranışları sergileyebileceğini göstermektedir (Kahneman ve Tversky, 1979). Örneğin, bireyler belirli bir kazancı garanti eden bir seçeneği, daha yüksek ancak belirsiz bir kazanç fırsatına kıyasla tercih etme eğilimindedir. Aynı şekilde, büyük kayıpların olduğu senaryolarda, bireyler daha yüksek risk almayı kabul edebilmektedir. Bu tür davranışsal eğilimler, finansal piyasalar, sağlık kararları ve işletme stratejileri gibi birçok alanda gözlemlenmektedir (Thaler, 1980).

KDS ve algoritmalar, organizasyonlara risk değerlendirme süreçlerinde yardımcı olmak için büyük ölçekli verileri analiz etmektedir. Ancak, bu sistemlerin risk algısını nasıl şekillendirdiği ve yönlendirdiği hala tartışmalı bir konudur. Algoritmalar, genellikle optimizasyon fonksiyonları aracılığıyla “en iyi” kararı belirlemeye çalışmakta, ancak karar vericilerin psikolojik ve duygusal faktörlerini göz önüne almamaları nedeniyle insan algısıyla uyumsuz sonuçlar üretebilmektedir (Rahwan vd., 2019).

3.2.2. Algoritmaların Risk Tercihlerini Şekillendirmesi

Algoritmalar, risk yönetimi süreçlerinde geniş veri kümelerini analiz ederek belirli kararlar için en uygun seçenekleri sunmaya çalışmaktadır. Örneğin, kredi tahsis algoritmaları, bir bireyin krediye uygun olup olmadığını geçmiş finansal davranışlara dayalı olarak değerlendirmektedir. Ancak bu süreç, algoritmaların risk tercihlerinin nasıl programlandığına bağlı olarak değişkenlik göstermektedir (Ustun ve Rudin, 2019). Algoritmaların risk tercihlerini şekillendirme yolları şu temel faktörlere dayanır:

3.2.2.1. Veri Seçimi ve Modelleme Süreçleri

Algoritmaların eğitim verisi, geçmişten gelen önyargıları içerebilmektedir. Örneğin, geçmişte belirli etnik grupların daha düşük kredi puanı aldığı bir veri seti, algoritmanın bu grupları riskli olarak sınıflandırmasına neden olabilmektedir (O'Neil, 2016).

3.2.2.2. Optimizasyon Fonksiyonları

Algoritmalar genellikle belirli bir hedefi optimize etmek üzere tasarlanmaktadır. Bir sigorta şirketinin algoritması, müşteri riskini minimize etmek için kayıpları önceden tahmin etmeye çalışırken, algoritmanın risk algısı geçmiş verilere dayalı olarak şekillenebilmekte ve mevcut ayrımcılığı sürdürmektedir (Mehrabi vd., 2021).

3.2.2.3. Riskten Kaçınma veya Aşırı Risk Alma Stratejileri

Beklenti Teorisi'ne göre, insanlar genellikle kayıplardan kaçınmak için daha güvenli seçeneklere yönelmektedirler. Algoritmalar, bu insan eğilimlerini modelleyerek karar önerileri sunabilir. Örneğin, borsa yatırım algoritmaları, yatırımcıların aşırı riskli işlemler yapmasını önlemek için belirli volatilité eşiklerini kullanarak kararlar alternatifleri oluşturabilirler (Chorafas, 2011). Ancak, algoritmaların kararlarını açıklanabilir kılmaması veya insan faktörünü göz ardı etmesi, karar vericilerin algoritmik önerilere güvenmesini zorlaştırmaktadır (Doshi-Velez ve Kim, 2017).

3.2.3. Algoritmik Yanlılığın Beklenti Teorisi Bağlamında Değerlendirilmesi

Beklenti Teorisi'nin öne sürdüğü temel kavramlardan biri, insanların kazançları ve kayıpları farklı şekillerde değerlendirmesidir. Algoritmalar, bu insan eğilimlerini modellemek üzere tasarlanabilir ancak eğitim verisindeki yanlılıklar, sistematik hatalara yol açmaya devam edecektir. Örneğin, yargı sistemlerinde kullanılan risk değerlendirme algoritmaları, sanıkların tekrar suç işleme olasılığını tahmin ederken belirli grupları daha yüksek risk kategorisine sokmaktadır. Angwin ve diğerleri (2016) tarafından yapılan bir araştırma, COMPAS adlı risk değerlendirme sisteminin Afro-Amerikan sanıkları daha yüksek riskli olarak sınıflandırdığını göstermiştir. Bu, Beklenti Teorisi açısından değerlendirildiğinde, algoritmanın kayıplara karşı aşırı duyarlılık göstererek belirli demografik grupları yanlış bir şekilde yüksek riskli olarak etiketlemesine neden olmuştur.

Benzer şekilde, sağlık sektöründe kullanılan tahmin modelleri, belirli hastalıkların teşhisinde belirli etnik grupları dezavantajlı hale getirmektedir.

Obermeyer ve diğerleri (2019) tarafından yapılan bir araştırma, sağlık algoritmalarının beyaz hastalar için daha fazla kaynak tahsis ederken, siyahi hastalara daha az öncelik verdiğini ortaya koymuştur. Bu tür yanlılıklar, Beklenti Teorisi bağlamında ele alındığında, algoritmaların belirli grupları yanlış bir şekilde “daha az öncelikli” olarak sınıflandırmasıyla ilişkilendirilmektedir.

Tüm bu örnekler, algoritmik karar sistemlerinin insan risk algısına uygun hale getirilmesi ve yanlılıkların önlenmesi için daha şeffaf ve etik tasarımlar geliştirilmesi gerektiğini göstermektedir.

4. Algoritmik Yanlılık, Etik Karar Verme ve FAT Framework

Algoritmik karar verme süreçleri, büyük veri ve YZ teknolojilerinin hızla yaygınlaşmasıyla birlikte toplumsal ve kurumsal hayatın önemli bir parçası haline gelmiştir. Ancak, bu sistemler veri yanlılığı, modelleme hataları ve karar destek süreçlerindeki sistematik önyargılar nedeniyle belirli grupları dezavantajlı hale getirebilmektedir (Barocas vd., 2019). Bu bağlamda, etik karar verme süreçleri, Fairness (Adalet), Accountability (Hesap Verebilirlik) ve Transparency (Şeffaflık) kavramları çerçevesinde incelenmektedir. FAT Framework olarak bilinen bu etik çerçeve, algoritmik sistemlerin adil, denetlenebilir ve şeffaf olması gerektiğini vurgulayan temel bir yaklaşım olarak kabul edilmektedir (Mittelstadt vd., 2016).

4.1. Adalet Kavramı ve Algoritmik Eşitlik

Algoritmik adalet, bir algoritmanın farklı bireyler veya gruplar arasında ayırım yapmadan adil kararlar üretmesi gerektiğini ifade etmektedir. Ancak, adalet kavramı farklı bağlamlarda değişiklik gösterebilir ve belirli bir algoritmanın ne kadar adil olduğu, uygulama alanına ve etik yaklaşıma göre farklı şekillerde yorumlanabilir (Dwork vd., 2012).

4.1.1. Algoritmik Kararlarda Adaletin Tanımı

Adalet kavramı, farklı perspektiflerden ele alınmaktadır. Sonuçsal adalet, sürece dayalı adalet ve fırsat eşitliği gibi farklı modeller, algoritmik kararların adaletini değerlendirmek için kullanılan yaklaşımlardır (Binns, 2018). Örneğin, işe alım süreçlerinde kullanılan bir makine öğrenimi modeli belirli bir cinsiyet veya etnik gruba daha az olumlu sonuç üretiyorsa, bu algoritma fırsat eşitliği ilkesine aykırı hareket ediyor denebilir (Raghavan vd., 2020). Aynı şekilde, kredi tahsis sistemlerinde belirli demografik grupların daha az kredi onayı alması, algoritmik adaletin ihlal edildiğini göstermektedir (Ustun ve Rudin, 2019).

4.1.2. Algoritmik Ayrımcılığı Önleme Stratejileri

Adil algoritmalar geliştirmek için kullanılan stratejilerden biri, önleyici modelleme ve veri setlerini dengeleme teknikleridir (Mehrabi vd., 2021). Algoritmaların eğitiminde kullanılan veri setlerinin daha çeşitli hale getirilmesi, belirli grupların dışlanmasını önlemek için önemli bir adımdır. Örneğin, Fairness Through Awareness ve Adversarial Debiasing gibi teknikler, algoritmaların belirli özellikleri (cinsiyet, etnik köken vb.) doğrudan veya dolaylı olarak kullanarak ayrımcılık yapmasını engellemeyi amaçlamaktadır (Zafar vd., 2017).

4.1.3. Algoritmik Eşitlik için Adil Modelleme Yöntemleri

Adaletin sağlanması için “parite tabanlı” ve “grup-farklılığı” yaklaşımları yaygın olarak kullanılmaktadır. Parite tabanlı yöntemler, tüm grupların aynı kararları almasını garanti etmeye çalışırken, grup-farklılığı yöntemleri, farklı grupların özel ihtiyaçlarına uygun şekilde algoritmaların tasarlanmasını hedeflemektedir (Hardt vd., 2016). Örneğin, sağlık sistemlerinde kullanılan tahmin algoritmalarının belirli bir etnik grubu yanlış teşhis etme oranı daha yüksekse, bu algoritmanın yeniden eğitilmesi ve çeşitli grupları daha dengeli bir şekilde temsil eden veri setleri ile test edilmesi gerekmektedir (Obermeyer vd., 2019).

4.2. Hesap Verebilirlik ve Algoritmaların Denetimi

Hesap verebilirlik, algoritmaların karar verme süreçlerinde açıklanabilir ve denetlenebilir olması gerektiğini ifade etmektedir. Algoritmaların çıktıları üzerinde denetim yapılmadığında, yanlış kararların tespit edilmesi ve düzeltilmesi zorlaşmaktadır (Rahwan vd., 2019).

4.2.1. Algoritmik Kararların Şeffaflığı ve Sorumluluk İlkeleri

Şeffaf olmayan algoritmaların neden belirli kararları verdiğini anlamak oldukça zor hale gelmektedir. Özellikle kara kutu problemi, algoritmaların iç işleyişinin belirsiz kalmasına neden olarak açıklanabilirlik ve güvenilirlik açısından önemli endişeler doğurmaktadır (Lipton, 2018). Bu durum, özellikle yüksek etkili karar süreçlerinde algoritmik hesap verebilirlik ihtiyacını daha da kritik hale getirmektedir.

Kurumsal düzeyde, algoritmik kararların yasal düzenlemelere uygun olması ve etik denetimlere tabi tutulması gerekmektedir. Örneğin, GDPR, bireylerin algoritmik kararlarla ilgili açıklama talep etmesine izin vermektedir.

4.2.2. Algoritmaların Açıklanabilirliği

AYZ yaklaşımları, algoritmaların verdiği kararları daha anlaşılır hale getirmeyi ve şeffaflığı artırmayı amaçlamaktadır. LIME (Local Interpretable Model-Agnostic Explanations) ve SHAP (Shapley Additive Explanations) gibi teknikler, modelin karar alma süreçlerini analiz ederek insan yorumuna açık hale getirmeye çalışmaktadır (Doshi-Velez ve Kim, 2017). Bu yaklaşımlar, özellikle yüksek etkili karar mekanizmalarında güvenilirlik ve hesap verebilirliği sağlamak için kritik bir rol oynamaktadır.

4.2.3. Kurumlarda Hesap Verebilirlik Mekanizmalarının Güçlendirilmesi

Hesap verebilirliği sağlamak ve algoritmik karar süreçlerinin güvenilirliğini artırmak için şirketlerin kapsamlı etik denetim mekanizmaları oluşturması gerekmektedir. Bu bağlamda, şeffaf veri yönetimi politikalarının geliştirilmesi, algoritmaların nasıl eğitildiği ve hangi verilerin kullanıldığı konusunda net standartlar belirlenmesini sağlamalıdır. Ayrıca, algoritmik kararların düzenli olarak denetlenmesini sağlayan bağımsız kurulların kurulması, bu süreçlerin objektif bir şekilde değerlendirilmesine ve olası yanlılıkların tespit edilmesine katkıda bulunacaktır. Bu tür önlemler, YZ ve veri odaklı sistemlerin adil, güvenilir ve etik standartlara uygun şekilde çalışmasını destekleyerek kurumsal sorumluluğun güçlendirilmesine yardımcı olmaktadır (Binns, 2018).

4.3. Şeffaflık ve Algoritmik Kararların Açıklanabilirliği

Şeffaflık, algoritmaların nasıl çalıştığını, hangi verileri kullandığını ve hangi kriterlere göre karar verdiğini açıklayabilme yeteneğidir. Şeffaf bir algoritma, karar alma süreçlerini anlaşılır ve denetlenebilir hale getirerek güvenilirliği artırırken, şeffaf olmayan algoritmalar ise yanlış kararlar üretebilir ve hesap verebilirliği azaltabilmektedir. Bu durum, özellikle kritik karar süreçlerinde, algoritmaların öngörülemez ve denetlenemez hale gelmesi nedeniyle etik ve yönetsel riskleri beraberinde getirmektedir (Mittelstadt vd., 2016).

4.3.1. Algoritmaların Kara Kutu Problemi

Kara kutu algoritmalar, karar alma süreçlerini dış dünyadan gizleyen ve iç işleyişi anlamamayan YZ modelleridir (Lipton, 2018). Bu tür algoritmalar, girdileri alarak belirli çıktılar üretse de bu kararların nasıl alındığına dair açıklamalar sunamaz. Özellikle kredi skorlama, işe alım, sağlık ve hukuk gibi kritik karar alma alanlarında, algoritmik şeffaflığın eksikliği büyük riskler doğurmaktadır. Yanlı veya hatalı kararlar, bireyler ve kurumlar için

ciddi ekonomik, sosyal ve hukuki sonuçlara yol açabilmektedir. Bu nedenle, algoritmik hesap verebilirliği sağlamak ve etik riskleri en aza indirmek için AYZ yaklaşımlarına ve düzenleyici denetim mekanizmalarına olan ihtiyaç giderek artmaktadır.

4.3.2. Algoritmik Karar Verme Süreçlerinde Şeffaflık Modelleri

Şeffaf algoritmalar, kullanıcıların sistemin nasıl çalıştığını, hangi verileri kullandığını ve ne tür çıktılar ürettiğini anlayabilmesini sağlayan yapılar olarak tanımlanmaktadır. Bu tür algoritmalar, karar alma süreçlerinin daha öngörülebilir ve denetlenebilir olmasını mümkün kılarak hesap verebilirlik, güvenilirlik ve etik uyumluluk açısından önemli avantajlar sunmaktadır. Bu bağlamda, White Box (Beyaz Kutu) Modelleme gibi yaklaşımlar, algoritmaların iç yapısını daha iyi anlamayı ve karar alma mekanizmalarını detaylı bir şekilde analiz etmeyi sağlamaktadır (Danks ve London, 2017). Beyaz kutu modeller, kullanılan değişkenlerin ağırlıklarını ve etkileşimlerini açıkça göstererek, karar süreçlerinin açıklanabilir ve denetlenebilir olmasına olanak tanımaktadır.

Şeffaflık modelleri arasında kural tabanlı sistemler, karar ağaçları, genişletilmiş lineer modeller ve AYZ yaklaşımları gibi yöntemler öne çıkmaktadır. Bu modeller, özellikle yüksek etkili sektörlerde yanlılığın azaltılması ve etik karar süreçlerinin desteklenmesi açısından kritik bir rol oynamaktadır.

4.3.3. Açıklanabilir Yapay Zekâ Çözümleri

AYZ çözümleri, YZ sistemlerinin karar alma süreçlerini anlamlandırmayı ve insan yorumuna açık hale getirmeyi amaçlamaktadır (Doshi-Velez ve Kim, 2017). Geleneksel kara kutu modellerin aksine, AYZ yaklaşımları algoritmaların nasıl çalıştığını, hangi verileri kullandığını ve kararlarını hangi faktörlere dayandığını şeffaf bir şekilde ortaya koyarak hesap verebilirliği artırmaktadır. Bu çözümler, özellikle sağlık, finans, hukuk ve insan kaynakları gibi yüksek etkili karar süreçlerinde modelin güvenilirliğini ve etik uyumluluğunu sağlamak açısından büyük önem taşımaktadır. AYZ teknikleri, modelin öngörülerini açıklamak, olası hataları belirlemek ve algoritmik yanlılığı en aza indirmek için geliştirilmektedir.

Öne çıkan AYZ teknikleri şunlardır:

- LIME: Modelin bireysel tahminlerini anlamlandırmak için yerel açıklamalar sunan bir yöntemdir (Ribeiro vd., 2016).

- SHAP: Modelin karar alma sürecinde hangi değişkenlerin ne ölçüde etkili olduğunu gösteren bir analiz yaklaşımıdır (Lundberg ve Lee, 2017).
- Karar Ağaçları ve Kurallı Modeller: Beyaz kutu modelleme kapsamında, karar mekanizmalarını daha şeffaf hale getiren yöntemlerdir (Quinlan, 1996).
- Karar Destek Araçları ve Görselleştirme Teknikleri: Kullanıcıların YZ modellerinin işleyişini daha iyi anlamasına yardımcı olan interaktif grafikler ve görselleştirme araçlarıdır (Molnar vd., 2020).

AYZ çözümlerinin yaygınlaştırılması, yalnızca algoritmik hesap verebilirliği artırmakla kalmayıp aynı zamanda kullanıcı güvenini de sağlamlaştırarak YZ'nin etik ve sürdürülebilir şekilde benimsenmesini desteklemektedir.

5. Algoritmik Yanlılığın Yönetimi ve Çözüm Önerileri

Algoritmik yanlılık, veri setlerindeki dengesizlikler, modelleme süreçlerindeki önyargılar ve karar destek mekanizmalarının tasarımındaki eksikliklerden kaynaklanmaktadır. Bu tür yanlılıklar, kredi skorlama, işe alım, sağlık teşhisleri ve kamu güvenliği gibi kritik alanlarda ciddi etik ve toplumsal sorunlara yol açabilmektedir (Barocas vd., 2019). Bu nedenle, algoritmik yanlılığın tespit edilmesi, yönetilmesi ve azaltılması için sistematik yaklaşımlar geliştirilmelidir. En etkili yöntemlerden biri, veri çeşitliliğini artırmak, model değerlendirme süreçlerini güçlendirmek ve etik regülasyonları uygulamaktır (Mehrabi vd., 2021). Bu bölümde, algoritmik yanlılığı azaltmaya yönelik stratejiler ele alınmaktadır.

5.1. Algoritmik Yanlılığı Azaltmak için Veri ve Model Optimizasyonu

Algoritmik sistemlerde yanlılığın önlenmesi için hem veri seviyesinde hem de modelleme süreçlerinde optimizasyon yapılması gerekmektedir. Veri tarafında, temsil edici ve dengeli veri setleri oluşturmak büyük önem taşımaktadır. Veri toplama aşamasında demografik çeşitlilik sağlanmalı ve azınlık gruplarını dışlamayan kapsamlı veri kümesi oluşturulmalıdır. Özellikle örnekleme hataları, eksik veri problemleri ve veri dağılımındaki dengesizlikler, algoritmaların sistematik yanlılık üretmesine neden olabilmektedir (Žliobaitė, 2017).

Modelleme sürecinde ise algoritmik yanlılığı ölçmek ve gidermek için çeşitli adil öğrenme teknikleri kullanılmaktadır. Bu teknikler, modelin

kararlarının farklı gruplar üzerinde adil olmasını sağlamayı amaçlamakta ve yanlılığın azaltılmasına yönelik yöntemler sunmaktadır.

Öne çıkan adil öğrenme ve yanlılık azaltma teknikleri arasında çeşitli yaklaşımlar bulunmaktadır. Önışleme yöntemleri, veri setlerindeki dengesizlikleri gidermek amacıyla örnekleme, ağırlıklandırma ve veri dönüştürme tekniklerinin uygulanmasını içermektedir (Kamiran ve Calders, 2012). Model sırasında teknikleri, adil öğrenmeyi sağlamak için düzenlenmiş kayıp fonksiyonları ve kısıtlamalar ekleyerek modelin belirli demografik gruplara yönelik ayrımcılık yapmasının önüne geçmeyi hedeflemektedir (Zafar vd., 2017). Son işleme yaklaşımları ise modelin tahmin sonuçlarını analiz ederek, belirli gruplara yönelik ayrımcılığı azaltan düzeltmelerin uygulanmasını sağlamaktadır (Hardt vd., 2016).

Bu optimizasyon süreçleri, YZ ve veri bilimi uygulamalarında etik karar almayı teşvik ederek hesap verebilirliği güçlendirmekte ve toplumsal etkileri minimize etmektedir.

5.1.1. Veri Çeşitliliği ve Temsiliyetin Artırılması

Algoritmik yanlılık çoğunlukla, belirli grupların yeterince temsil edilmediği veya yanlış temsil edildiği veri setlerinden kaynaklanmaktadır. Eğer bir makine öğrenimi modeli yalnızca belirli bir demografik gruba ait verilerle eğitilirse, model diğer gruplar için düşük doğrulukta veya sistematik hatalar üreten kararlar almaktadır (Buolamwini ve Gebru, 2018).

Temsiliyetin artırılması için önerilen stratejiler arasında çeşitli yaklaşımlar öne çıkmaktadır. Dengeli ve çeşitlendirilmiş veri setleri kullanımı, tüm demografik grupları ve değişken özellikleri kapsayacak şekilde veri setlerinin oluşturulmasını gerektirmektedir. Örneğin, yüz tanıma sistemlerinde farklı etnik grupların, cinsiyetlerin ve yaş kategorilerinin eşit oranda temsil edilmesi, modelin adaletli ve tarafsız çalışmasını sağlamaktadır.

Sentetik veri üretimi ve veri artırma teknikleri, eksik veya az temsil edilen gruplara yönelik veri dengesini sağlamak amacıyla kullanılmaktadır. Özellikle sağlık sektöründe, nadir görülen hastalıkları temsil eden sentetik hasta verileri üretilerek daha kapsayıcı modellerin geliştirilmesi mümkün olmaktadır (Obermeyer vd., 2019).

Veri temizleme ve yanlı verilerin tespit edilmesi ise eğitim verilerinin dikkatlice analiz edilerek, önyargılı, yanlış veya eksik bilgiler içeren veri noktalarının ayıklanmasını içermektedir (Dwork vd., 2012). Bu süreç, modelin öğrenme aşamasında yanlı kararlar üretmesini önlemeye yardımcı olmaktadır.

5.1.2. Model Değerlendirme ve Yanlılık Testleri

Makine öğrenimi ve YZ modelleri, geliştirme aşamasında belirli testlerden geçirilerek yanlı olup olmadıkları analiz edilmesi gereklidir. Bunun için çeşitli algoritmik yanlılık ölçütleri ve adil modelleme teknikleri kullanılmaktadır (Mehrabi vd., 2021).

Yanlılık testleri için kullanılan bazı yöntemler:

5.1.2.1. Demografik Parite Testi

Modelin her demografik grup için aynı olasılıkla olumlu kararlar verip vermediğini ölçmektedir. Örneğin, bir kredi tahsis modelinin kadın ve erkek başvurularına eşit oranlarda onay verip vermediği kontrol edilmesi gösterilebilir (Hardt, Price ve Srebro, 2016).

5.1.2.2. Adil Fırsat Eşitliği

Modelin, farklı gruplar için eşit doğrulukta kararlar ürettiğini doğrulamaya çalışmaktadır. Sağlık teşhis modellerinde farklı etnik gruplara yönelik tahmin başarısı ölçülerek belirli grupların yanlış teşhis oranlarının daha yüksek olup olmadığı analiz edilmesi örnek olarak gösterilebilir (Obermeyer vd., 2019).

5.1.2.3. Yanlılık Giderme Teknikleri

Model eğitimi sırasında Fairness Constraints ve Adversarial Debiasing gibi teknikler kullanılarak belirli değişkenlerin (cinsiyet, etnik köken vb.) model tarafından ayrımcı bir şekilde öğrenmesi engellenebilir (Zafar vd., 2017).

5.1.2.4. Modelin Hassasiyet ve Önyargı Değerlendirmesi

Algoritmanın kararlarının açıklanabilirliğini artırmak için SHAP ve LIME gibi AYZ yöntemleri uygulanabilir (Doshi-Velez ve Kim, 2017).

Yanlılık testleri, yalnızca model geliştirme aşamasında değil, aynı zamanda model dağıtımına alındıktan sonra düzenli olarak yapılmalıdır. Zira, modeller zaman içinde değişen veri akışları nedeniyle ilk eğitim dönemine kıyasla yeni yanlılıklar geliştirebilmektedir (Mittelstadt vd., 2016).

5.1.3. Regülasyonlar ve Standartlar ile Algoritmik Yanlılığın Önlenmesi

Algoritmik sistemlerin etik ve adil çalışmasını sağlamak için çeşitli yasal düzenlemeler ve endüstri standartları geliştirilmiştir. Avrupa Birliği, ABD ve

diğer ülkeler, algoritmik karar verme süreçlerinin etik kurallara uygun olması ve hesap verebilirliği artırılması için belirli yasal çerçeveler oluşturmuştur.

Öne çıkan regülasyonlar ve standartlar, algoritmik sistemlerin etik, adil ve hesap verebilir şekilde tasarlanmasını sağlamak amacıyla geliştirilmiştir. GDPR, Avrupa Birliği tarafından yürürlüğe konulmuş olup bireylerin algoritmik karar süreçlerinde nasıl değerlendirildiklerini açıklama hakkına sahip olmalarını gerektirmektedir. Özellikle, otomatik kararların neden alındığı ve bireylerin bu kararlara itiraz etme hakkına sahip olduğu açıkça belirtilmelidir (Wachter vd., 2017).

ABD YZ Etik Rehberleri, ABD'de çeşitli kurumlar tarafından YZ sistemlerinin etik sorumluluklarını belirlemek ve algoritmik karar süreçlerinde hesap verebilirliği artırmak amacıyla geliştirilmiştir (Danks ve London, 2017).

IEEE Etik YZ Standartları, YZ sistemlerinin adil, güvenilir ve açıklanabilir olması gerektiğini vurgulayan uluslararası bir çerçeve sunmaktadır (Schiff vd., 2020). Bunun yanı sıra, büyük teknoloji şirketleri de etik YZ ve yanlılık azaltma konularında iç politikalar geliştirerek, modellerini düzenli olarak denetlemektedir. Örneğin, Google, Microsoft ve IBM gibi şirketler, şeffaflık ve hesap verebilirlik ilkelerini gözeterek algoritmik sistemlerini sürekli olarak gözden geçirmektedir (Binns, 2018).

Bu regülasyonlar ve standartlar, algoritmik yanlılığın önlenmesi ve adil YZ uygulamalarının geliştirilmesi için temel bir çerçeve sunmaktadır.

5.2. Geleceğe Yönelik Stratejik Öneriler ve Akademik Araştırma Alanları

Algoritmik yanlılık, mevcut YZ ve KDS'nin en büyük zorluklarından biri olmaya devam etmektedir. Veri kaynaklarının çeşitlendirilmesi, modelleme süreçlerinin adil hale getirilmesi ve hesap verebilir yapılar oluşturulması, algoritmik sistemlerin daha etik ve güvenilir olmasını sağlayacak kritik faktörlerdir (Barocas vd., 2019).

Bu bağlamda, gelecekteki akademik araştırmalar ve uygulamalar, yanlılık önleme tekniklerinin geliştirilmesi, kurumsal düzeyde algoritmik denetim stratejilerinin uygulanması ve YBS'de etik standartların belirlenmesi gibi temel alanlara odaklanmalıdır (Mittelstadt vd., 2016).

5.2.1. Algoritmik Yanlılığın Önlenmesinde Yeni Nesil Yaklaşımlar

Geleneksel algoritmik yanlılık giderme yöntemleri, çoğunlukla veri temizleme, model optimizasyonu ve denetimli öğrenme süreçlerinin yeniden

tasarlanması gibi tekniklere odaklanmaktadır. Ancak, yeni nesil yaklaşımlar, yanlılığı sadece mevcut veriler ve modellerle sınırlandırmayıp, daha geniş yapısal çözümler geliştirmeyi hedeflemektedir.

Bu bağlamda, YZ sistemlerinde adil ve tarafsız karar alma süreçlerini desteklemek amacıyla geliştirilen yeni nesil yaklaşımlar öne çıkmaktadır. Adversarial Debiasing, derin öğrenme modellerinde adil öğrenme mekanizmaları kullanarak yanlı verilerin etkisini azaltmayı hedefleyen bir yöntemdir (Zafar vd., 2017). Bu teknik, modelin belirli değişkenlere karşı ayrımcı sonuçlar üretip üretmediğini tespit ederek öğrenme sürecini optimize etmektedir.

Merkezsiz öğrenme yaklaşımları, verilerin merkezi bir noktada toplanmadan, dağıtılmış ağlar üzerinden işlenmesini sağlayarak veri güvenliği ve anonimliği artırırken, yanlılık risklerini minimize etmeyi amaçlamaktadır (McMahan vd., 2017).

Nedensel çıkarım modelleri, algoritmaların yanlı kararlar üretip üretmediğini anlamak için geleneksel korelasyon tabanlı analizlerin ötesine geçerek neden-sonuç ilişkilerini analiz eden yaklaşımlar sunmaktadır (Pearl, 2009). Bu sayede, sistemlerin öğrenme süreçlerinde önyargı kaynaklarının daha etkin bir şekilde belirlenmesi ve giderilmesi sağlanmaktadır.

Son olarak, evrimsel algoritmalar ve genetik programlama, YZ sistemlerinin zaman içinde önyargılı kararları tespit edip düzeltebilecek şekilde evrimsel süreçlerle geliştirilmesini mümkün kılmaktadır. Bu yöntem, uzun vadeli adaletli modelleme açısından önemli bir adım olarak değerlendirilmektedir (Bäck vd., 1997).

Gelecekteki araştırmalar, bu yeni nesil yaklaşımların performansını artırarak daha güvenilir, adil ve etik YZ sistemleri geliştirmeye odaklanacağı düşünülmektedir.

5.2.2. İşletmeler ve Kamu Kurumları için Algoritmik Denetim Stratejileri

Kurumsal düzeyde YZ ve veri analitiği sistemlerinin yaygınlaşması, şirketler ve kamu kurumları için algoritmik denetim stratejilerinin geliştirilmesini zorunlu hale getirmiştir (Rahwan vd., 2019). Denetim mekanizmalarının oluşturulması hem iç hem de dış denetim süreçleriyle algoritmaların adil, hesap verebilir ve şeffaf çalışmasını sağlamayı hedeflemektedir.

Kurumsal ve kamu düzeyinde algoritmik denetim mekanizmalarının güçlendirilmesi, YZ sistemlerinin hesap verebilirliğini ve şeffaflığını

artırmak açısından kritik öneme sahiptir. Bu doğrultuda, çeşitli stratejiler önerilmektedir.

Bağımsız algoritmik denetim kurulları, büyük ölçekli organizasyonlarda YZ sistemlerinin düzenli etik denetimlerden geçirilmesini sağlayarak KDS'lerin hesap verebilirliğini artırmaktadır (Binns, 2018).

Açıklanabilirlik ve izlenebilirlik kriterleri, algoritmaların nasıl çalıştığını daha iyi anlamak için AYZ tekniklerinin kullanımını gerektirmektedir. Bu sayede, yöneticiler ve denetçiler, YZ çıktılarının hangi faktörlere dayandığını analiz edebilmekte ve olası yanlışlık kaynaklarını belirleyebilmektedir (Doshi-Velez ve Kim, 2017).

Algoritmik etki değerlendirme çerçevesi, Avrupa Komisyonu tarafından önerilmiş olup, algoritmaların bireyler ve toplum üzerindeki potansiyel etkilerini sistematik bir analiz süreciyle değerlendirmeyi amaçlamaktadır (Mittelstadt vd., 2016).

Son olarak, iç denetim ve şeffaflık politikaları, şirketlerin veri toplama süreçlerinden model geliştirme aşamasına kadar algoritmalarını bağımsız birimler tarafından değerlendirmeye açmasını ve çalışanların algoritmik yanlışlık riskleri konusunda bilinçlendirilmesini sağlamaktadır (Danks ve London, 2017).

Bu denetim mekanizmaları, YZ tabanlı karar süreçlerinin toplumsal eşitlik ilkeleriyle uyumlu olmasını sağlamak için önemli bir araç olacaktır.

5.2.3. Yönetim Bilişim Sistemlerinde Algoritmik Etik Standartlarının Geliştirilmesi

YBS, organizasyonların bilgi yönetimi ve karar alma süreçlerini optimize eden sistemlerdir. Ancak, YBS süreçlerinde kullanılan YZ ve makine öğrenimi modelleri etik riskler taşıyabilir. Bu nedenle, YBS bağlamında algoritmik etik standartlarının oluşturulması, sistemlerin adil, şeffaf ve hesap verebilir bir şekilde çalışmasını garanti etmek için kritik bir gerekliliktir (Shapiro ve Varian, 2013).

Algoritmik sistemlerin etik çerçevede geliştirilmesi ve uygulanması, adalet, hesap verebilirlik, şeffaflık ve veri koruma ilkeleri doğrultusunda şekillendirilmelidir. Bu bağlamda, önerilen etik standartlar şu şekilde özetlenebilir:

Adalet, YBS kapsamında geliştirilen YZ ve veri analitiği çözümlerinin tüm kullanıcı gruplarına eşit hizmet sunmasını ve belirli demografik gruplara karşı ayrımcılık yapmamasını gerektirmektedir (Hardt, Price ve Srebro, 2016).

Hesap verebilirlik, kurumların algoritmik sistemlerin ürettiği hatalardan doğrudan sorumlu olmasını ve hatalı kararların düzeltilmesine yönelik mekanizmalar geliştirmesini zorunlu kılmaktadır (Rahwan vd., 2019).

Şeffaflık, kullanıcıların sistemlerin nasıl çalıştığını ve karar süreçlerinde hangi verilerin kullanıldığını anlayabilmesini sağlayarak algoritmik süreçlerin güvenilirliğini artırmaktadır (Lipton, 2018).

Gizlilik ve veri koruma, algoritmaların etik standartlara uygun şekilde geliştirilmesini ve kullanılan kişisel verilerin uluslararası regülasyonlara, özellikle GDPR gibi düzenlemelere uygun olarak saklanmasını gerektirmektedir (Wachter, Mittelstadt ve Floridi, 2017).

Son olarak, sürekli izleme ve güncelleme, algoritmaların zaman içinde değişen veri kaynaklarına uyum sağlamasını ve etik ilkelerin dinamik bir şekilde gözden geçirilerek güncellenmesini içermektedir (Obermeyer vd., 2019).

Bu etik standartlar, YBS perspektifinden algoritmik yanlılığın önlenmesi ve YZ tabanlı karar destek mekanizmalarının daha güvenilir hale getirilmesi açısından büyük önem taşımaktadır.

6. Sonuç ve Değerlendirme

Algoritmik yanlılık, YZ tabanlı sistemlerin gelişimi ile birlikte kurumsal, toplumsal ve etik açıdan önemli riskler taşıyan bir olgu olarak karşımıza çıkmaktadır. KDS, makine öğrenimi modelleri ve büyük veri analitiği süreçleri, işletmelerin ve kamu kurumlarının karar alma süreçlerini büyük ölçüde etkilemektedir. Ancak, bu sistemlerin kullandığı veriler ve algoritmaların tasarım süreçleri, tarafsızlık ilkesine aykırı önyargıları içerebilmekte ve bazı gruplar için sistematik ayrımcılık yaratabilmektedir (Barocas vd., 2019).

6.1. Algoritmik Yanlılığın Kurumsal, Toplumsal ve Etik Boyutları

Algoritmik sistemler, iş dünyasında karar alma süreçlerini hızlandırarak operasyonel verimliliği artırmaktadır. Ancak, işe alım, terfi, kredi tahsis ve fiyatlandırma gibi kritik kararlar YZ tarafından alındığında, bu sistemlerin adaletli olup olmadığı önemli bir tartışma konusu haline gelmektedir. Örneğin, insan kaynakları departmanlarında kullanılan makine öğrenimi algoritmaları, geçmiş işe alım verilerine dayalı olarak adayları değerlendirdiğinde, kadınlar veya belirli etnik gruplar gibi dezavantajlı kesimlere karşı ayrımcılık yapabilmektedir (Raghavan vd., 2020).

Bu tür algoritmik yanlılıklar yalnızca iş dünyasıyla sınırlı kalmayıp toplumsal düzeyde de ciddi adaletsizliklere yol açabilmektedir. Örneğin,

kredi skorlama algoritmalarının belirli gelir gruplarına daha az kredi verme eğiliminde olması, ekonomik eşitsizlikleri derinleştirebilmektedir (Ustun ve Rudin, 2019). Benzer şekilde, sağlık hizmetlerinde kullanılan YZ sistemleri, belirli etnik gruplara mensup bireyleri yanlış sınıflandırarak tıbbi teşhis süreçlerinde hatalara neden olabilmektedir (Obermeyer vd., 2019).

Bu bağlamda, YZ sistemleri geliştirilirken adalet, hesap verebilirlik ve şeffaflık ilkelerinin gözetilmesi büyük önem taşımaktadır. FAT Framework, algoritmaların etik olarak nasıl geliştirileceğine dair önemli bir çerçeve sunmaktadır (Mittelstadt vd., 2016). Algoritmik sistemlerin etik olmayan kararlar üretmemesi için denetim mekanizmalarının geliştirilmesi ve bu sistemlerin daha açıklanabilir hale getirilmesi gerekmektedir (Doshi-Velez ve Kim, 2017).

6.2. Algoritmik Karar Verme ve Yönetim Bilişim Sistemlerinde Uygulamalar

YBS, organizasyonların karar alma süreçlerini optimize eden bilgi sistemlerinden oluşmaktadır. Ancak, YBS içerisinde kullanılan algoritmaların yanlılık içermemesi ve karar süreçlerinin adil bir şekilde yürütülmesi büyük önem taşımaktadır.

İZ, büyük veri analitiği ve tahmine dayalı modelleme süreçleri, organizasyonların stratejik kararlarını yönlendiren önemli araçlar haline gelmiştir. Ancak, bu süreçlerin kurumsal yapılar içinde yanlı kararlar üretme riski göz ardı edilmemelidir. Örneğin, kredi risk analizinde kullanılan algoritmalar, belirli sosyo-ekonomik grupları dezavantajlı hale getirebilmekte ve ekonomik eşitsizlikleri artırabilmektedir (Dwork vd., 2012).

Benzer şekilde, kamu yönetimi ve politika oluşturma süreçlerinde de algoritmaların adil kararlar ürettiğini garanti altına almak gerekmektedir. Hukuk sisteminde kullanılan risk değerlendirme algoritmaları, suç tekrarını öngörmeye hatalı sonuçlar üretebilmekte ve belirli grupları yanlış bir şekilde yüksek risk kategorisine sokabilmektedir (Angwin vd., 2016).

Bu bağlamda, YBS içerisinde kullanılan KDS'nin etik ve hesap verebilir olabilmesi için çeşitli hususların göz önünde bulundurulması gerekmektedir. Öncelikle, adil ve tarafsız veri setlerinin kullanılması, algoritmaların yanlı kararlar üretme riskini azaltacaktır. Bunun yanı sıra, algoritmaların bağımsız denetimlerden geçirilmesi, olası hataların ve önyargıların tespit edilmesine katkı sağlayacaktır. Ayrıca, şeffaflık ilkelerinin benimsenmesi ve karar süreçlerinin anlaşılabilir hale getirilmesi, kullanıcıların algoritmik sistemlere olan güvenini artıracaktır. Son olarak, GDPR gibi uluslararası regülasyonlara

uygun sistemlerin geliştirilmesi, veri gizliliğini ve etik standartları koruma açısından kritik bir gereklilik olarak öne çıkmaktadır.

Bu unsurlar dikkate alındığında, YBS'nin kurumsal karar süreçlerinde etik yapay zeka kullanımını destekleyen sağlam bir çerçeve oluşturduğu söylenebilir.

6.3. Etik Yapay Zekâ ve Algoritmik Kararlar İçin Gelecek Perspektifleri

Algoritmik karar sistemlerinin yaygınlaşmasıyla birlikte, etik YZ geliştirme süreçleri ve algoritmik yanlılığı önleme stratejileri daha fazla önem kazanmaktadır. YZ tabanlı sistemlerin giderek daha fazla sektörde kullanılması, bu sistemlerin adil, hesap verebilir ve şeffaf hale getirilmesini zorunlu kılmaktadır. Ancak, mevcut algoritmaların birçok alanda yanlı çıktılar ürettiği ve bu durumun bireyler ve topluluklar için olumsuz sonuçlar doğurduğu bilinmektedir. Bu nedenle, gelecekteki YZ uygulamalarında etik çerçevenin güçlendirilmesi, algoritmaların daha kapsayıcı hale getirilmesi ve karar mekanizmalarının açıklanabilirliği konusunda ilerleme kaydedilmesi gerekmektedir.

YZ sistemlerinin etik ve sosyal sorumluluk çerçevesinde geliştirilmesi, yalnızca teknik çözümlerle sınırlı kalmamalıdır. Adil modelleme teknikleri, algoritmaların eğitim süreçlerinde kullanılan verilerin homojen ve temsili olmasını sağlamak için geliştirilen yöntemlerdir. Örneğin, federated learning gibi merkezsiz öğrenme yöntemleri, veri güvenliğini korurken, yanlılık risklerini en aza indirmek için yeni bir çerçeve sunmaktadır. Aynı zamanda, nedensel çıkarım modelleri, YZ'nin belirli demografik gruplara yönelik haksız önyargılar geliştirmesini önlemek için veri analizine yeni bir perspektif kazandırmaktadır. Bununla birlikte, YZ sistemlerinin hesap verebilir olması için algoritmik denetim mekanizmalarının oluşturulması gerekmektedir. Büyük ölçekli organizasyonlar, algoritmik süreçlerin düzenli olarak incelenmesini sağlayacak bağımsız denetim kurulları oluşturmalıdır. Bu tür denetim süreçleri, yalnızca algoritmaların doğru çalışmasını garanti altına almakla kalmaz, aynı zamanda algoritmaların ürettiği öngörülerin toplumsal ve bireysel etkilerini de analiz etme imkânı sunar. Aynı şekilde, algoritmaların karar süreçlerini daha anlaşılır hale getirmek için AYZ yaklaşımlarının geliştirilmesi büyük önem taşımaktadır. Algoritmaların hangi kriterleri kullanarak belirli kararları verdiğinin anlaşılır olması hem kullanıcılar hem de düzenleyici kurumlar için güvenilirliği artıracaktır.

YZ uygulamalarında etik standartların belirlenmesi ve geliştirilmesi, yalnızca şirketlerin veya akademik araştırmaların sorumluluğunda değildir.

YZ regülasyonlarının güçlendirilmesi, hükümetlerin ve uluslararası kuruluşların sorumluluk alanına giren kritik bir konudur. GDPR gibi düzenlemeler, algoritmik karar verme süreçlerinde bireylerin haklarını korumayı hedeflemektedir. Ancak, bu tür regülasyonların küresel ölçekte daha kapsamlı hale getirilmesi gerekmektedir. YZ kullanımına yönelik yasal çerçevelerin genişletilmesi, özellikle bireylerin algoritmik kararlar hakkında itiraz etme ve kararların nasıl verildiğini anlama haklarını koruma altına alınmalıdır.

Özetle, etik YZ sistemleri geliştirmek, adil veri kullanımı, hesap verebilir algoritmik süreçler ve şeffaf karar mekanizmaları oluşturmakla mümkündür. Teknolojik ilerlemeler, YZ'nin insan hayatına entegrasyonunu hızlandırırken, etik sorunların göz ardı edilmemesi gerekmektedir. Akademik araştırmalar ve kurumsal politikalar, algoritmaların insan haklarına saygılı ve toplumsal değerlerle uyumlu şekilde geliştirilmesine katkı sağlamalıdır. Gelecekte, YZ sistemlerinin etik ilkeler doğrultusunda çalışmasını sağlamak için sistematik denetim, adil modelleme teknikleri ve yasal düzenlemeler giderek daha fazla önem kazanacaktır. Algoritmik karar verme süreçlerinin toplumsal adalet ve etik kurallar çerçevesinde geliştirilmesi, YZ'nin sürdürülebilir ve güvenilir bir şekilde ilerlemesi için kaçınılmaz bir gerekliliktir.

Kaynakça

- Angwin, J., Larson, J., Mattu, S., ve Kirchner, L. (2022). Machine bias. In *Ethics of data and analytics* (pp. 254-264). Auerbach Publications.
- Augier, M. (2001). Simon says: Bounded rationality matters: Introduction and interview. *Journal of Management Inquiry*, 10(3), 268-275. <https://doi.org/10.1177/1056492601103010>
- Back, T., Hammel, U., ve Schwefel, H. P. (1997). Evolutionary computation: Comments on the history and current state. *IEEE transactions on Evolutionary Computation*, 1(1), 3-17. <https://doi.org/10.1109/4235.585888>
- Bakshy, E., Messing, S., ve Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132. <https://doi.org/10.1126/science.aaa1160>
- Barocas, S., ve Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, 104, 671. <https://doi.org/10.15779/Z38BG31>
- Barocas, S., Hardt, M., ve Narayanan, A. (2023). *Fairness and machine learning: Limitations and Opportunities*. MIT press.
- Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy ve Technology*, 31(4), 543-556. <https://doi.org/10.1007/s13347-017-0263-5>
- Brynjolfsson, E., ve McAfee, Andrew. (2017). Artificial intelligence, for real. *Harvard Business Review*, 1(1), 1-31.
- Buolamwini, J., ve Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77-91). PMLR.
- Caliskan, A., Bryson, J. J., ve Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186. <https://doi.org/10.1126/science.aal4230>
- Chorafas, D. N. (2011). *Stress testing for risk control under Basel II*. Elsevier.
- Çam, S. (2024). Empowering marketing intelligence via text analytics. In *Marketing innovation strategies and consumer behavior* (pp. 31-57). IGI Global. <https://doi.org/10.4018/979-8-3693-4195-7.ch002>
- Danks, D., ve London, A. J. (2017, August). Algorithmic bias in autonomous systems. In *Ijcai* (Vol. 17, No. 2017, pp. 4691-4697).
- Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters (2018).
- Davenport, T., ve Harris, J. (2017). *Competing on analytics: Updated, with a new introduction: The new science of winning*. Harvard Business Press.
- Doshi-Velez, F., ve Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. <https://doi.org/10.48550/arXiv.1702.08608>

- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., ve Zemel, R. (2012, January). Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference* (pp. 214-226).
- Erbey, A. (2024). Yapay zekâ. İçinde M. F. Karaca (Ed.), *Yönetim Bilişim Sistemleri alanında uygulama, kuram ve kavramlar* (ss. 102-116). Duvar Yayınları.
- Yılmaz, E. O., (2022). Uzaktan eğitimde çevrimiçi destekli yüz yüze gözetimli sınav uygulaması: Bir model önerisi. İçinde P. Ünüvar (Ed.), *Eğitim Bilimleri Alanında Yeni Trendler 1* (ss. 91-104), İzmir: Duvar Yayınları.
- Gigerenzer, G., ve Selten, R. (2001). Rethinking rationality. Bounded rationality: The adaptive toolbox (pp. 1–12). The MIT Press.
- Gündüz, C., ve Polat, H. (2021). Turkish sign language recognition based on multistream data fusion. *Turkish Journal of Electrical Engineering and Computer Sciences*, 29(2), 1171-1186.
- Hardt, M., Price, E., ve Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, 29.
- Kahneman, D. ve Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 363-391.
- Kamiran, F., ve Calders, T. (2012). Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33(1), 1-33. <https://doi.org/10.1007/s10115-011-0463-8>
- Kim, B., Kim, H., Kim, K., Kim, S., ve Kim, J. (2019). Learning not to learn: Training deep neural networks with biased data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9012-9020).
- Laudon, K. C., ve Laudon, J. P. (2017). *Essentials of management information systems*. Pearson.
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., ve Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy ve Technology*, 31(4), 611-627. <https://doi.org/10.1007/s13347-017-0279-x>
- Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31-57.
- Lundberg, S. M., ve Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- McMahan, B., Moore, E., Ramage, D., Hampson, S., ve y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.

- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., ve Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1-35. <https://doi.org/10.1145/3457607>
- Menon, A. K., ve Williamson, R. C. (2018, January). The cost of fairness in binary classification. In *Conference on Fairness, Accountability and Transparency* (pp. 107-118). PMLR.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., ve Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data ve Society*, 3(2). <https://doi.org/10.1177/205395171667967>
- Molnar, C., Casalicchio, G., ve Bischl, B. (2020, September). Interpretable machine learning—a brief history, state-of-the-art and challenges. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 417-431). Cham: Springer International Publishing.
- Obermeyer, Z., Powers, B., Vogeli, C., ve Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. <https://doi.org/10.1126/science.aax2342>
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, Newyork.
- Pearl, J. (2009). Causal inference in statistics: An overview. <https://doi.org/10.1214/09-SS057>
- Power, D. J., Sharda, R., ve Burstein, F. (2015). Decision support systems. Volume 7. Management information systems. *Cooper C.L. Wiley Encyclopedia of Management*. New York: John Wiley ve Sons, Ltd, 1-11.
- Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys (CSUR)*, 28(1), 71-72.
- Raghavan, M., Barocas, S., Kleinberg, J., ve Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 469-481). <https://doi.org/10.1145/3351095.3372828>
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ve Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477-486. <https://doi.org/10.1038/s41586-019-1138-y>
- Raisch, S., ve Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192-210. <https://doi.org/10.5465/amr.2018.0072>
- Ribeiro, M. T., Singh, S., ve Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*. <https://doi.org/10.48550/arXiv.1606.05386>
- Schiff, D., Ayesh, A., Musikanski, L., ve Havens, J. C. (2020, October). IEEE 7010: A new standard for assessing the well-being implications of artifi-

- cial intelligence. In 2020 IEEE international conference on systems, man, and cybernetics (SMC) (pp. 2746-2753). IEEE.
- Shapiro, C., ve Varian, H. R. (1999). *Information rules: A strategic guide to the network economy*. Harvard Business Press.
- Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). MIT press.
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of economic behavior ve organization*, 1(1), 39-60. [https://doi.org/10.1016/0167-2681\(80\)90051-7](https://doi.org/10.1016/0167-2681(80)90051-7)
- Tschantz, M. C., Datta, A., Datta, A., ve Wing, J. M. (2015, July). A methodology for information flow experiments. In *2015 IEEE 28th Computer Security Foundations Symposium* (pp. 554-568). IEEE. <https://doi.org/10.1109/CSE.2015.40>
- Tuna, M. F., ve Görmez, Y. (2024). Evrimsel sinir ağları tabanlı derin öğrenme yöntemiyle müşteri şikayetlerinin sınıflandırılması. *Bingöl Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 8(1), 31-46. <https://doi.org/10.33399/biibfad.1362160>
- Ustun, B., ve Rudin, C. (2019). Learning optimized risk scores. *Journal of Machine Learning Research*, 20(150), 1-75.
- Wachter, S., Mittelstadt, B., ve Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International data privacy law*, 7(2), 76-99. <https://doi.org/10.1093/idpl/ix005>
- Zafar, M. B., Valera, I., Rogriguez, M. G., ve Gummadi, K. P. (2017, April). Fairness constraints: Mechanisms for fair classification. In *Artificial intelligence and statistics* (pp. 962-970). PMLR.
- Žliobaitė, I. (2017). Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*, 31(4), 1060-1089. <https://doi.org/10.1007/s10618-017-0506-1>

