MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENİMİ TABANLI KARAR DESTEK SİSTEMLERİNDE ALGORİTMİK ÖNYARGI VE EŞİTSİZLİK ÜZERİNE BİR ARAŞTIRMA: KREDİ DEĞERLENDİRME VAKA ANALİZİ

YÜKSEK LİSANS TEZİ

Mert Burak Dinçman

Enformatik Ana Bilim Dalı

Bilgisayar Ortamında Sanat ve Tasarım

Tez Danışmanı: Doç. Dr. Nazım Ziya Perdahçı

İstanbul, Haziran 2023

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna uygun olarak hazırladığım bu tez çalışmasında;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel etik kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ücret karşılığı başka kişilere yazdırmadığımı (dikte etme dışında), uygulamalarımı yaptırmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

Mert Burak Dinçman

Bu tez çalışmasının tamamlanmasında desteklerini esirgemeyen, cesaretlendiren, yönlendiren ve tez çalışmamın başarılı bir şekilde tamamlanmasında katkısı olan Tez Danışmanım Sayın Nazım Ziya Perdahçı'ya, aileme ve arkadaşlarıma teşekkür ederim.

Ek olarak Sayın Nazım Ziya Perdahçı tez çalışmamın her aşamasında bana rehberlik ederek, yol gösterdi ve fikirlerime önem vererek beni motive etti. Yönlendirmesi, bilgi birikimi ve deneyimleri sayesinde tez çalışmamın kalitesini artırdı ve tamamlanmasında büyük katkısı oldu. Ayrıca, sabırlı tutumu ve her zaman yanımda olması benim için çok değerliydi.

Mert Burak Dinçman

MAKİNE ÖĞRENİMİ TABANLI KARAR DESTEK SİSTEMLERİNDE ALGORİTMİK ÖNYARGI VE EŞİTSİZLİK ÜZERİNE BİR ARAŞTIRMA: KREDİ DEĞERLENDİRME VAKA ANALİZİ

ÖZET

Bu çalışmada, insanların günlük yaşamlarında sıklıkla karşılaştığı Yapay Zekâ, makine öğrenmesi ve algoritmaların artan maruziyeti ele alınmaktadır. Bununla birlikte, bu teknolojilerin kullanımının yaygınlaşmasıyla birlikte, algoritmik önyargı ve makine öğrenmesi adaletsizliği gibi yeni kavramların ortaya çıktığına dikkat çekilmektedir. Makine öğrenmesinin kısıtlı bir alanda aldığı kararların fark yaratmasına rağmen, kapsam alanının artmasıyla birlikte algoritmaların etki alanı dar kapsamlı olmaktan çıkabildiği ve toplumsal ayrışmalara neden olabileceği ifade edilmektedir. Bu nedenle, kredi karar süreçlerinde kullanılan makine öğrenmesi algoritmalarının önyargı veya eşitsizliğe neden olup olmadığı Alman Kredi verisi üzerinde test edilmiştir. Test sonuçları istatistiksel sonuçlar ve eşitsizlik metrikleri yardımıyla yorumlanmış ve makine öğrenmesi algoritmalarının kredi süreçlerinde adaletsizliğe sebep olduğu belirlenmiştir. Algoritmik önyargıların, Yapay Zekânın insanlara kıyasla geniş kapsamlı öğrenme ve anlamlandırma yeteneğinin olmaması, Yapay Zekânın adaptasyon yeteneğinin yetersiz kalmasının yanı sıra duygusal zekâsı gibi yeteneklerinin eksikliğinden kaynaklanabileceği tartışılmıştır. Ayrıca, makine öğrenmesinin adaptasyon yeteneğinin olmaması değişimlerde Yapay Zekânın genel Zekâya göre uyumsuz kalmasına neden olduğu vurgulanmaktadır. Bu çalışma, yönetim bilimleri açısından bir optimizasyon problemi olarak karar destek sistemi için açıklanabilir Yapay Zekânın önemini ve çözüm gücü de vurgulamaktadır. Algoritmik önyargı ve eşitsizlik metrikleri gibi konular, yönetim bilimleri alanında karar alıcıların Yapay Zekâ teknolojileri kullanımında dikkate almaları gereken önemli hususlardır. Ayrıca, Yapay Zekânın duygusal boyutunun eksikliği, algoritmik önyargıları ortaya çıkarabileceğinden, bu teknolojilerin geliştirilmesinde duygusal zekâ bileşenlerinin dikkate alınması önemlidir. Bu çalışma, Yapay Zekâ teknolojilerinin etik, hukuk ve toplumsal boyutlarına da dikkat çekerek, Yapay Zekâ kullanımının toplumsal yararına katkı sağlamayı amaçlamaktadır.

***Anahtar Kelimeler**: Yapay Zekâ (AI), Açıklanabilir Yapay Zekâ (XAI), Makine Öğrenmesi, Algoritmik Önyargı, Eşitsizlik Metrikleri, Alman Kredi Verisi, Yönetim Bilimleri, Optimizasyon Problemi, Karar Destek Sistemi*

A RESEARCH ON ALGORITHMIC BIAS AND FAIRNESS IN MACHINE LEARNING-BASED SUPPORT DECISION SYSTEMS: A CASE ANALYSIS OF CREDIT EVALUATION

ABSTRACT

In this study, the increasing exposure to artificial intelligence, machine learning, and algorithms that individuals commonly encounter in their daily lives is addressed. However, with the proliferation of the use of these technologies, new concepts such as algorithmic bias and machine learning unfairness have emerged. While the decisions taken by machine learning in a limited area can make a difference, it is stated that with the expansion of its scope, the impact of algorithms can go beyond narrow domains and lead to societal divisions. Therefore, machine learning algorithms used in credit decision-making processes were tested for bias or inequality on German credit data. Test results were interpreted using statistical results and inequality metrics, and it was determined that machine learning algorithms cause unfairness in credit processes. It has been discussed that algorithmic biases may arise from the lack of extensive learning and interpretation abilities of artificial intelligence compared to humans, as well as insufficient adaptability and missing abilities such as emotional intelligence. In addition, it is emphasized that the lack of adaptability of machine learning causes artificial intelligence to be incompatible with general intelligence in changes. This study highlights the importance and solution power of explainable artificial intelligence for decision support systems as an optimization problem from the perspective of management sciences. Issues such as algorithmic bias and inequality metrics are important considerations for decision-makers in the use of artificial intelligence technologies in management sciences. Furthermore, since the lack of emotional dimension of artificial intelligence may lead to algorithmic biases, it is important to consider emotional intelligence components in the development of these technologies. This study aims to contribute to the societal benefit of the use of artificial intelligence by also drawing attention to its ethical, legal, and societal dimensions.

Keywords: *Artificial Intelligence (AI), Explainable Artificial Intelligence (XAI), Machine Learning, Algorithmic Fairness, Algorithmic Bias, Inequality Metrics, German Credit Data*

İÇİNDEKİLER

Sayfa

KISALTMALAR

AKVS	: Alman Kredi Veri Seti
CBD DO	: Cumhurbaşkanlığı Dijital Dönüşüm Ofisi
DN	: Doğru Negatif
DP	: Doğru Pozitif
FE	: Fark Etkisi
FEF	: Fırsat Eşitliği Farkı
GB	: Gigabyte
İPF	: İstatiksel Parite Farkı
KDS	: Karar Destek Sistemi
LRM	: Lojistik Regresyon Modeli
MÖ	: Makine Öğrenmesi
OOF	: Ortalama Oran Farkı
TDK	: Türk Dil Kurumu
TÜBA	: Türkiye Bilimler Akademisi
TBTS	: Türkçe Bilim Terimleri Sözlüğü
XAI	: Açıklanabilir Yapay Zekâ
YN	: Yanlış Negatif
YÖKTEZ	: Yükseköğretim Kurulu Tez Merkezi
YP	: Yanlış Pozitif
YZ	: Yapay Zekâ
ZB	: Zettabyte

ÇİZELGE LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

MAKİNE ÖĞRENİMİ TABANLI KARAR DESTEK SİSTEMLERİNDE ALGORİTMİK ÖNYARGI VE EŞİTSİZLİK ÜZERİNE BİR ARAŞTIRMA: KREDİ DEĞERLENDİRME VAKA ANALİZİ

ÖZET

Bu çalışmada, insanların günlük yaşamlarında sıklıkla karşılaştığı Yapay Zekâ, makine öğrenmesi ve algoritmaların artan maruziyeti ele alınmaktadır. Bununla birlikte, bu teknolojilerin kullanımının yaygınlaşmasıyla birlikte, algoritmik önyargı ve makine öğrenmesi adaletsizliği gibi yeni kavramların ortaya çıktığına dikkat çekilmektedir. Makine öğrenmesinin kısıtlı bir alanda aldığı kararlar ile fark yaratmasına rağmen, kapsam alanının artmasıyla birlikte algoritmaların etki alanı dar kapsamlı olmaktan çıkabildiği ve toplumsal ayrışmalara neden olabileceği ifade edilmektedir. Bu nedenle, kredi karar süreçlerinde kullanılan makine öğrenmesi algoritmalarının önyargı veya eşitsizliğe neden olup olmadığı Alman Kredi verisi üzerinde test edilmiştir. Test sonuçları istatistiksel sonuçlar ve eşitsizlik metrikleri yardımıyla yorumlanmış ve makine öğrenmesi algoritmalarının kredi süreçlerinde adaletsizliğe sebep olduğu belirlenmiştir. Algoritmik önyargıların, Yapay Zekânın insanlara kıyasla geniş kapsamlı öğrenme ve anlamlandırma yeteneğinin olmaması, Yapay Zekânın adaptasyon yeteneğinin yetersiz kalmasının yanı sıra duygusal zekâsı gibi yeteneklerinin eksikliğinden kaynaklanabileceği tartışılmıştır. Ayrıca, makine öğrenmesinin adaptasyon yeteneğinin olmaması, değişimlerde Yapay Zekânın genel Zekâya göre uyumsuz kalmasına neden olduğu vurgulanmaktadır. Bu çalışma, yönetim bilimleri açısından bir optimizasyon problemi olarak karar destek sistemi için açıklanabilir Yapay Zekânın önemini ve çözüm gücü de vurgulamaktadır. Algoritmik önyargı ve eşitsizlik metrikleri gibi konular, yönetim bilimleri alanında karar alıcıların Yapay Zekâ teknolojileri kullanımında dikkate almaları gereken önemli hususlardır. Ayrıca, Yapay Zekânın duygusal boyutunun eksikliği, algoritmik önyargıları ortaya çıkarabileceğinden, bu teknolojilerin geliştirilmesinde duygusal zekâ bileşenlerinin dikkate alınması önemlidir. Bu çalışma, Yapay Zekâ teknolojilerinin etik, hukuk ve toplumsal boyutlarına da dikkat çekerek, Yapay Zekâ kullanımının toplumsal yararına katkı sağlamayı amaçlamaktadır.

***Anahtar Kelimeler**: Yapay Zekâ (AI), Açıklanabilir Yapay Zekâ (XAI), Makine Öğrenmesi, Algoritmik Önyargı, Eşitsizlik Metrikleri, Alman Kredi Verisi, Yönetim Bilimleri, Optimizasyon Problemi, Karar Destek Sistemi*

A RESEARCH ON ALGORITHMIC BIAS AND FAIRNESS IN MACHINE LEARNING-BASED SUPPORT DECISION SYSTEMS: A CASE ANALYSIS OF CREDIT EVALUATION

ABSTRACT

In this study, the increasing exposure to artificial intelligence, machine learning, and algorithms that individuals commonly encounter in their daily lives is addressed. However, with the proliferation of the use of these technologies, new concepts such as algorithmic bias and machine learning unfairness have emerged. While the decisions taken by machine learning in a limited area can make a difference, it is stated that with the expansion of its scope, the impact of algorithms can go beyond narrow domains and lead to societal divisions. Therefore, machine learning algorithms used in credit decision-making processes were tested for bias or inequality on German credit data. Test results were interpreted using statistical results and inequality metrics, and it was determined that machine learning algorithms cause unfairness in credit processes. It has been discussed that algorithmic biases may arise from the lack of extensive learning and interpretation abilities of artificial intelligence compared to humans, as well as insufficient adaptability and missing abilities such as emotional intelligence. In addition, it is emphasized that the lack of adaptability of machine learning causes artificial intelligence to be incompatible with general intelligence in changes. This study highlights the importance and solution power of explainable artificial intelligence for decision support systems as an optimization problem from the perspective of management sciences. Issues such as algorithmic bias and inequality metrics are important considerations for decision-makers in the use of artificial intelligence technologies in management sciences. Furthermore, since the lack of emotional dimension of artificial intelligence may lead to algorithmic biases, it is important to consider emotional intelligence components in the development of these technologies. This study aims to contribute to the societal benefit of the use of artificial intelligence by also drawing attention to its ethical, legal, and societal dimensions.

Keywords: *Artificial Intelligence (AI), Explainable Artificial Intelligence (XAI), Machine Learning, Algorithmic Fairness, Algorithmic Bias, Inequality Metrics, German Credit Data*

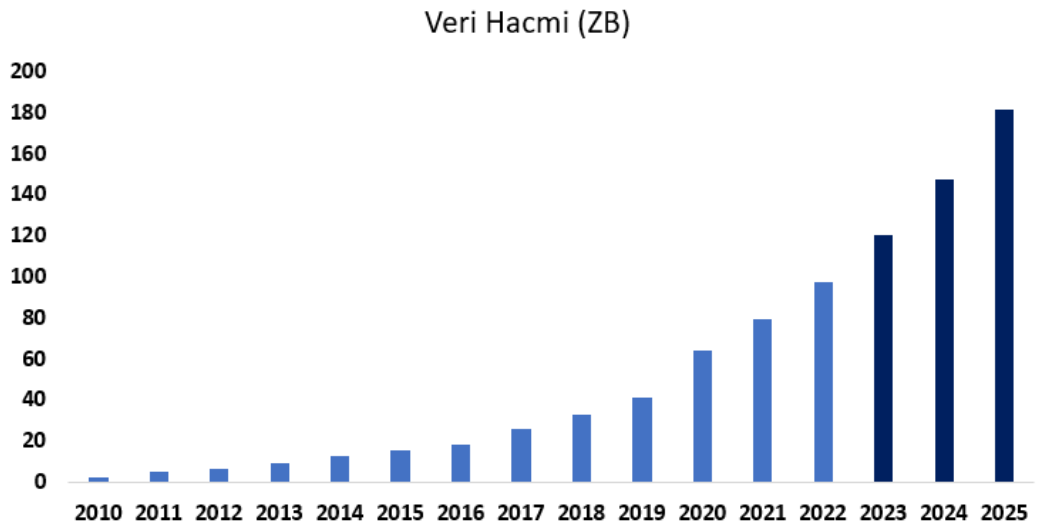
1. GİRİŞ

Yapay Zekâ (YZ) uygulamalarını, görev ve sorumlulukların insan odaklı operasyonel sürece bağlı kalmadan geliştirilmesini sağlayan teknolojik çözümler olarak tanımlayabiliriz. YZ uygulamaları, Makine Öğrenmesi (MÖ) teknolojilerini kullanarak iş yüklerini azaltmayı ve verimli şekilde çözümlenmesini hedefler. YZ ve MÖ kavramları son zamanlarda popüler olarak kullanılmakta olsa bile temeli 1950 yıllarına kadar dayanmaktadır. YZ ve MÖ kavramları ilk olarak 1950'li yıllarda John McCarthy, Alan Turing, Allen Newell ve Herbert A. Simon tarafından ele alınmıştır. YZ terimini ilk kullananlar arasında olan John McCarthy göre YZ, insan Zekâsının bilgisayarlar tarafından taklit edilmesi anlamına gelir [1]. Diğer bir önemli isim olan Alan Turing göre ise, makinelerin insanlarla kıyaslandığında dışarıdan ayırt edilemeyecek şekilde davranmasını YZ ve MÖ olarak tanımlamıştır [2]. Bu tanım aynı zamanda Turing Testi'nin temelini oluşturur [2]. 1958 yılında Allen Newell ve Herbert A. Simon insan dilini bilgisayar diline çevirmeyi hedeflemiş ve "Genel Problem Çözücü" adındaki program tasarımı ile problem çözebilen YZ sistemi geliştirmişlerdir [3]. Aynı zamanda ülkemizde 1959 yılında Ord. Prof Dr. Cahit Arf tarafından "Makine düşünebilir mi ve nasıl düşünebilir?" fikri ortaya atılmıştır [4].

YZ ve MÖ kavramları hakkındaki çalışmalar çok 1950 yıllarına kadar dayanmakta olmakla birlikte son yıllarda popülerlik kazanmasının ana sebepleri arasında veri ve artan veri hacmi kavramı yer alabilir. Veri kavramını farklı açılardan ele almak mümkündür. Chen vd. (2012) veriyi yapısal ve yapısal olmayan bilgi olarak tanımlamışlardır [1]. MIT tarafından yayınlanan diğer bir çalışmaya göre, Borgman (2015) veriyi en sade şekilde değer ve anlam taşıyan parça olarak tanımlamıştır [2]. Bu iki tanımlamadan yola çıkarak veriyi; toplanabilen, işlenebilen ve kullanılabilen form halindeki bir yapı taşına benzetebiliriz. Bu sayede veri farklı mimari oluşumların doğmasına neden olabilir. Artan veri hacmi ile yeni oluşumların aynı YZ ve MÖ uygulamalarında olduğu popüler hale gelmesini bekleyebiliriz.

Hilbert M., (2011) veri miktarının dünya genelinde hızla artış gösterdiğini ve gelişen teknoloji ile daha hızlı artacağını söylemiştir [3]. Artan veri hacmini anlamlandırabilmek için Statista tarafından geçmiş ve geleceği kapsayan araştırmaya göz atabiliriz [4]. Yapılan araştırmaya göre Dünya genelinde üretilen veri miktarının sırasıyla 2010 yılını için 2 zettabyte (ZB) (1 ZB = 10^12 GB, 1 GB = 10^9 byte), 2013 yılı için 9 zettabyte, 2016 yılı için 18 zettabyte ve 2020 yılı için 64 zettabyte olarak ölçülmesi ve 2025 yılı için 181 zettabyte olarak tahmin edilmesi Yapay Zekâ ve Makine Öğrenmesi uygulamalarının gelecekteki öneminin daha fazla artacağını göstermektedir. McKinsey tarafından büyük verinin önemini anlatmak için yayınlanan çalışmaya göre veri hacminin genişlemesi beraberin hem fırsatları hem de riskleri getirdiğini vurgulamıştır [5].

Çizelge 1.1'de görüldüğü üzere yüksek hızla artan veri sayısının, doğru şekilde ve uygun yöntemlerle analiz edilmesi gerekmektedir. Doğru şekilde yapılmayan çalışmalar fırsatların krizlere veya zorluklara dönüşmesine neden olabilir. Bu durum insanlık için önem ve sorumluluk arz etmektedir.



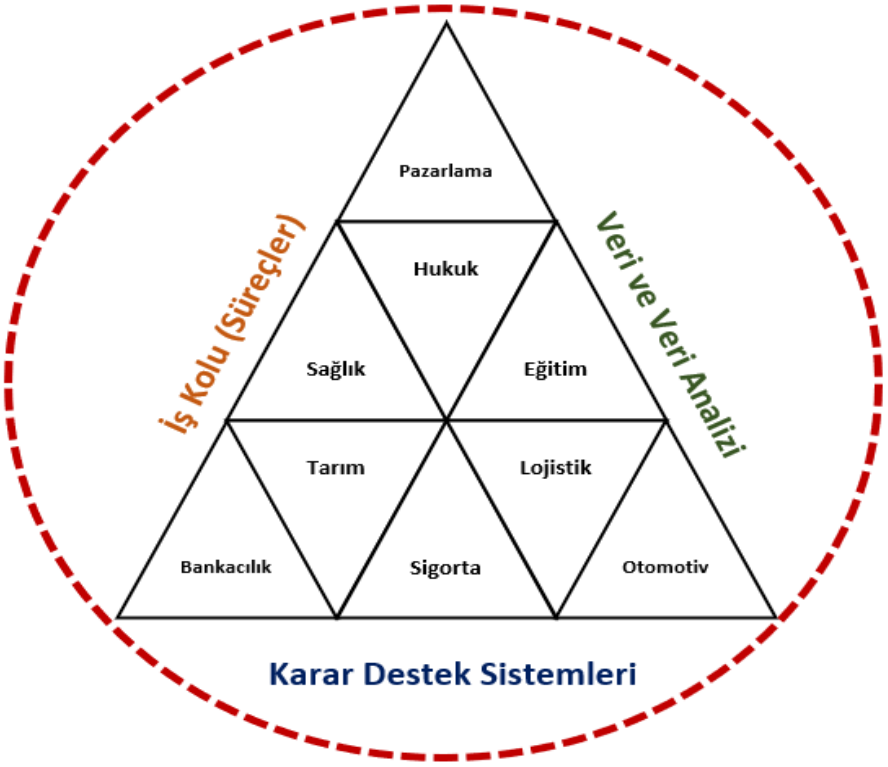
Çizelge 1.1 : Gelişen Veri Hacmi

Büyük veri ve analitiğinde yaşanan teknolojik gelişmeler, YZ ve MÖ uygulamalarının gelişmesini hızlandırmıştır.

Veri analizi çalışmalarında kullanılan YZ ve MÖ araçları, veri setlerini hızlı bir şekilde işleyebilme yetenekleri sayesinde birçok sektörde kritik role sahip olabilir.

Bu araçların arkasında yer alan algoritmaların kullanım sayısı ve süresinin artması, insan hayatındaki kapsamını da genişletmektedir. Bu teknolojiler, işletmelerin daha hızlı kararlar almasına yardımcı olabilirken aldığı kararlar neticesinde elinde olmayan sebeplerden dolayı farklı sonuçlara neden olabilir. YZ ve MÖ uygulamaları, şekil 1.1'de yer alan örnek sektörler gibi (tıp, tarım, eğitim ve bankacılık vb.) birçok endüstride de kullanılabilmektedir.

Veriyi günlük yaşamın enerji kaynağı olarak düşünecek olursak, algoritmaları enerji üreten nükleer santrallere benzetebiliriz. Nükleer santrallerin doğru kullanılmaması insanlık için ciddi sonuçlara neden olabilirken, YZ ve MÖ araçları ve arkasındaki algoritmaların doğru kullanılmaması şekil 1.1'de gösterilen insan hayatını kapsayan karar destek sistemlerinde değişimlerin yaşanmasına ve olumsuz sonuçların ortaya çıkmasına neden olabilmektedir. Karar destek sistemlerinde kullanılan YZ ve MÖ uygulamaları Yönetim Bilimlerinde verimlilik artışı vb. kolaylıklar getireceği gibi yaşanacak problemler nedeniyle zorluk çıkarması ve sosyal eşitsizliğe neden olması beklenmektedir. YZ ve MÖ uygulamalarının insan hayatına doğrudan müdahil olan farklı sektörlerde yaşanabilecek olası problem ve eşitsizlikleri göstermek adına yönetim bilimlerindeki konumlanması şekil 1.1'de gösterilmiştir.



Şekil 1.1 : YZ ve MÖ YB Konumlanması

Yönetim bilimi, işletmelerin etkin yönetilmesi ve doğru kararların alınmasını hedefler ve sonuçları iyileştirmek için çeşitli araçları kullanabilir. Bu kapsamda YZ ve MÖ araçlarının yönetim biliminde aktif kullanılması ve önemli bir role sahip muhtemeldir. YZ ve MÖ, karar süreçlerinde yöneticilerin kararlarını analizlerle desteklenmesinde, alınan kararlar için öngörü yapılmasında ve iş süreçlerini optimize edilmesinde kullanılabilir. İş dünyasının önde gelen kuruluşları ile yapılan anket ve değerlendirmeler sonrası üst yönetimin yönetim bilimlerinde YZ ve MÖ konumlandırmaya başladığı görülmektedir.

Bu kapsamda yapılan bazı araştırmalara aşağıda yer verilmiştir.

- PriceWaterCoopers (PwC) denetim ve danışmanlık şirketi tarafından yapılan araştırmaya göre [5] iş dünyasındaki yöneticilerin %72'sinin YZ ve MÖ uygulamalarının iş süreçlerinde avantaj sağladığını,

- MeMSQL yazılım şirketi tarafından yapılan araştırmaya göre [6] iş dünyası katılımcılarının %61'i YZ ve MÖ uygulamalarını gelecek yıllar sistemlerine entegre edeceğini,
- Adobe medya yazılım şirketinin yaptığı araştırmaya göre [7] ise şirketlerin %47'sinin YZ stratejisine sahip olduğunu,
- Gartner teknoloji araştırma ve danışmanlık şirketi tarafından yapılan araştırmaya göre [8], işletmelerin %37'si YZ ve MÖ teknolojileri kullandığını ve bu oranın 2022 yılına kadar %43'e yükselmesini beklediğini,
- IBM teknoloji ve bilgi işleme şirketi tarafından yaptığı araştırmaya göre [9], işletmelerin %94'ü YZ ve MÖ teknolojilerinin işletme stratejilerinde önemli bir rol oynayacağını,
- Dell teknoloji şirketi tarafından yapılan araştırmaya göre [10], işletmelerin %91'i YZ teknolojilerinin işletme içinde verimlilik artışı sağlayacağını,
- Infosys bilişim teknolojileri danışmanlık şirketi tarafından yapılan araştırmaya göre ise [11], işletmelerin %53'ü YZ ve MÖ teknolojilerinin işletmelerinin rekabet gücünü artıracağını söylemiştir.

Yukarıdaki araştırmalardan da anlaşılacağı gibi, YZ ve MÖ araçları yönetim biliminde giderek daha önemli bir role sahip olmaktadır. Bu araçlar, işletmelerin verimliliklerini artırmak, iş süreçlerini optimize etmek, doğru kararlar almak ve rekabet avantajı sağlamak için kullanılabilir. Bu nedenle, yöneticilerin YZ ve MÖ araçlarını bilmeleri ve kullanmaları, işletme performansını artırmak için önemli bir faktördür. Ayrıca, YZ ve MÖ araçlarının kullanımı, işletmelerin geleceğe yönelik stratejilerinin belirlenmesinde de yardımcı olabilir.

Şekil 1.1 ve iş dünyası tarafından yapılan araştırmalardan anlaşılacağı gibi algoritmaların karar destek sistemleri beraber iş süreçlerinde görev almaya başladığı görülmektedir. YZ ve MÖ uygulamaları karar destek sistemlerini dışardan destekleyebildiği gibi dijital inovasyon çıktıları sayesinde yönetim bilişim sürecine içerinden de katkı sağlayabilmektedir. Bu kapsamda kullanılan araçlar aşağıda özetlenmiştir.

- Sesli asistanlar: Sözlü taleplere yanıt veren bir Yapay Zekâ sistemidir. Kullanıcılara yardım sağlamayı hedefler.

- Otonom araçlar: Algoritmaları iletişimi sayesinde nesneler arası etkileşimi hedefler.
- Sohbet Robotu: Yazılı taleplere yanıt veren bir Yapay Zekâ sistemidir. Müşterileri ve kullanıcıların taleplerini karşılamayı hedefler.
- Robotlar: Hareket edebilen Yapay Zekâ sistemidir. Mekanik robotlar insan emek gücünün verimli kullanılması hedefler.

Yönetim bilişimde YZ uygulamalarının yaygınlığının artmasıyla MÖ teknolojisinin kullanımı önemli bir hal almıştır. MÖ sağladığı kolaylıklar ile başta karar destek sistemleri olmak üzere farklı süreç ve yönetişim yapılarında katma değer sağlayabilmektedir.

MÖ algoritmaları en basit anlatımla, mevcut verilerden çeşitli öğrenme teknikleri ile farklı görev ve süreçleri operasyonel iş yükü yaratmadan gerçekleştirilebilir. Bununla birlikte, MÖ uygulamaları avantaj sağladığı gibi öğrenme sürecinde yaşayacağı aksaklıklar nedeniyle dezavantaja da neden olabilir. MÖ nedeniyle yaşanabilecek dezavantajlar karar süreçlerinde eşitsizliklere sebep olabilir. Yaşanan bu durum, yönetim biliminde sıklıkla karşılaşılan optimizasyon problemine dikkat çekebilir [12] [13]. Optimizasyon problemlerini, mevcut kısıtlar altında ilgili hedefe en iyi çözümü arayan problem olarak özetleyebiliriz. Bu tanımı matematik dünyasında belirlenen fonksiyonu eldeki değişkenlerle minimize veya maksimize etme problemi olarak ifade edebiliriz [14] [15] [16] [17]. Optimizasyon teorisinin birçok disiplinde karşımıza çıkması muhtemeldir. Bu duruma en güzel örnek İktisat bilimi için yapılan "kısıtlı kaynaklarla maksimum fayda" tanımı gösterilebilir [12] [13].

YZ ve MÖ uygulamalarının işletmelerdeki en temel amacının karar süreçlerinde optimizasyon sağlamak olduğunu düşünecek olursak beraberinde eşitsizlik gibi problemleri getirmesi veya taşıması normal olarak karşılanabilir.

Karar destek süreçlerindeki sistemsel veya tasarımsal kaynaklı olumsuzluklar algoritma önyargısına sebep olmaktadır. Algoritma önyargısı, algoritmaların adaletsizlik veya eşitsizlik yaratması olarak bilinmektedir. YZ ve MÖ kullanımının artmasıyla ortaya çıkan algoritmik önyargıların çoğalması muhtemeldir.

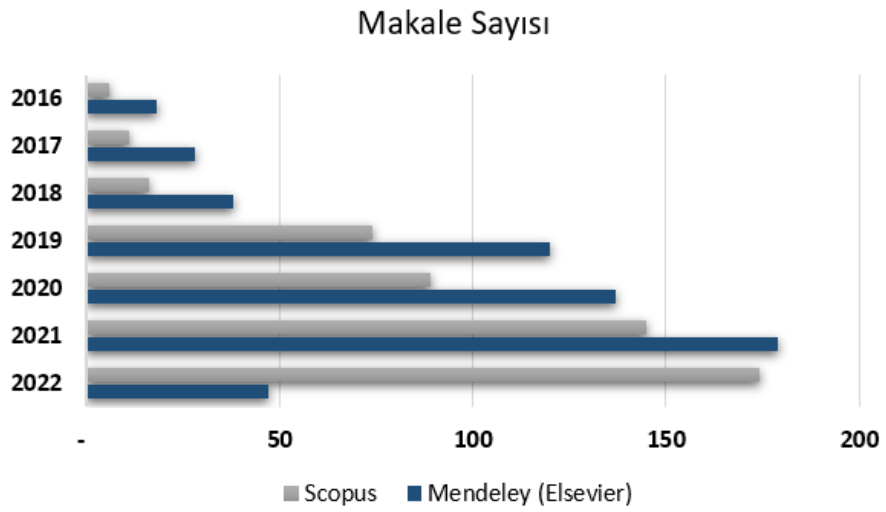
MÖ sebebiyle ortaya çıkacak eşitsizlikleri anlamlandırabilmek için MÖ sürecini baştan uca anlamak ve yorumlayabilmek gerekmektedir. Baştan uca çözümü optimizasyon problemindeki yaklaşım gibi öncelikle girdileri ele alarak daha sonra denkleme odaklanarak çözüm üretmek gerekebilir.

Bu kapsamda sırasıyla MÖ gelişmesine neden olan Yapay Zekâ uygulamaları ve Yapay Zekâ uygulamaların yaygınlığının artmasına neden olabilecek veri kavramlarını tanımlamak ve anlamlandırmak faydalı olabilir. Aynı zamanda bu optimizasyon sorunun çözümüne literatürde yine bir zekâ önerisi olan açıklanabilir Yapay Zekâ (XAI) kavramı çözüm aranmaktadır.

Açıklanabilir Yapay Zekâ (XAI), algoritmalarının kararları ve sonuçları nasıl aldığını kullanıcılara açıklayabilen ve örneklendirebilen bir yaklaşım olarak görülebilir [18]. Bu yaklaşım sayesinde, algoritmik önyargılar önlenebilir ve adil karar destek süreçleri oluşturulabilir. Karar destek süreçlerindeki sistem veya tasarım kaynaklı olumsuzluklar algoritma önyargısına sebep olmaktadır. Algoritma önyargısı, algoritmaların adaletsizlik veya eşitsizlik yaratması olarak bilinmektedir. YZ ve MÖ kullanımının artmasıyla ortaya çıkan algoritmik önyargıların çoğalması muhtemeldir.

Çizelge 1.2'de görüldüğü gibi algoritmik önyargı konusunda yapılan araştırma sayısının günden güne artması YZ ve MÖ süreçlerinde toplumsal farkındalığın arttığına örnek gösterilebilir.

Küresel ölçekte Scopus [19] ve Mendeley [20] üzerinde "Algorithmic Bias" anahtar kelimesi ile yapılan sorgulamalar sonucunda yıllar itibarıyla toplam 2022 yılı için 221 makale, 2021 yılı için 324, 2020 yılı için 226, 2019 yılı için 194, 2018 yılı için ise 54 makaleye ulaşılmaktadır. Çizelge 1.2'de görüldüğü gibi algoritmik önyargı çalışmalarının yakın zamanda ön plana çıktığı ve son 5 yılda artarak popülerlik kazandığını söylemek mümkündür. Yükseköğretim Kurulu Ulusal Tez Merkezi (YÖKTEZ) üzerinden [21] "Algoritmik Önyargı" vb. anahtar kelimeler ile yapılan sorgulamalar sonucunda güncel olarak akademik çalışmalara ulaşılmaması ulusal literatürdeki ihtiyacı göstermektedir.



Çizelge 1.2 : Makale Sayısı

Bu çalışma uygulamalı olarak karar destek sisteminin alt kümesi olan kredi karar süreçlerinde makine öğrenmesi algoritmaları ile yaşanabilecek sosyal eşitsizlikleri algoritmik önyargı kapsamında anlamlandırmayı hedeflemektedir. Önceki bölümlerde belirtildiği gibi kredi kararı sürecinde yaşanacak algoritmik önyargının devamında fırsat eşitsizliği, refah aktarımı vb. sonuçlara sebep olması muhtemeldir. Bu kapsamda çalışmanın ilerleyen bölümünde kredi karar süreçlerinde kullanılan YZ ve MÖ araçlarında algoritmik önyargı yoktur hipotezine, kredi kararı destek sürecinde kullanılan öğrenmesi algoritmalarını kullanarak cevap aramayı hedeflemektedir. Karşılaştırma sonucunda ortaya çıkan kararların sosyal eşitsizliğe neden olup olmadığı incelenecektir. Bu çalışma YZ, MÖ, algoritma, algoritmik önyargı kavramlarının ulusal ölçekte literatüre alternatif tanımlama ve ölçüm olarak somut içerik üretmeyi hedeflemektedir.

2. KAVRAMSAL ÇERÇEVE

2.1 Yapay Zekâ

Günümüzde YZ terminolojisi ile sıklıkla karşılaşılmaktadır. Karşılaşılan YZ kavramının kullanımı içinde bulduğu durumun şartlarına göre farklılık ve/veya benzerlik göstermesi muhtemeldir.

Türk Dil Kurumu'na (TDK) göre [22] "doğadaki örneklerine benzetilerek insan eliyle yapılmış anlamına gelen" yapay ve yine TDK'ya göre [22] "insanın düşünme, akıl yürütme, objektif gerçekleri algılama, yargılama ve sonuç çıkarma yeteneklerinin tamamı" anlamına gelen Zekâ sözcüklerinin birleşiminden oluşan YZ teriminin her iki kelimesinin odağında insanın olduğu açıkça görülmektedir.

Türkçe Bilim Terimleri Sözlüğü'ne (TBTS) göre [23] YZ terimi iki farklı disiplin olan sosyoloji ve mühendislik için ayrı ayrı tanımlanmıştır. YZ terimi sosyoloji disiplini altında "Bilgisayar mühendisliği, sinirbilim, felsefe, ruhbilim, robot bilimi ve dilbilim gibi birçok alanı içine alan ve algı, akıl yürütme, düşünme, öğrenme, kavrama, sezgi ve tasarlama gibi insan zekâsına özgü davranışlar sergileyen bilgisayar yazılımı, robot tasarımı gibi konuları inceleyen bilimsel alan ve bu biçimde ortaya çıkan ürün." olarak tanımlanırken; mühendislik disiplini altında YZ "Çoğu zaman algoritma biçiminde tanımlanamayan, buluşsal yöntemlerle otomatik öğrenme yöntemlerinden yararlanan ve doğal dil anlama, söz analiz, örüntü tanıma gibi algısal ya da bilişsel süreçlerle ilgili bilgisayar modelleri geliştiren araştırma alanı." olarak tanımlanmıştır.

Ayrıca resmi otorite tarafından (T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi) YZ terimi "Bilgisayarın veya bilgisayar kontrollü robotun, genellikle akıllı varlıklarla ilişkili görevleri yerine getirme yeteneği" olarak tanımlanmıştır [24]. Bu tanımı alt parçalarına ayırdığımız zaman Yapay Zekânın yetenek vurgusunun ön plana çıktığı görülmektedir.

Kurumların YZ tanımlamaları aşağıda Çizelge 2.1.1'de özetlenmiştir.

Çizelge 2.1.1: Kurumların YZ Tanımları ve Çıkarımlar

Kurum	Tanımlanan Kelime ve Terim	Mevcut Tanımın Özeti	Çıkarım Yapılan Anahtar Kelimeler
TDK	Yapay	İnsan eliyle yapılmış veya üretilmiş	İnsan, Ürün
TDK	Zekâ	İnsanın düşünme, algılama sonuç çıkarma vb. yeteneklerinin tamamı	İnsan, Algı, Çıkarım
TÜBA	Yapay Zekâ	İnsan özgü davranışlar sergileyen yazılım, ürün ve süreçler	İnsan, Süreç, Davranış
CBDDO	Yapay Zekâ	Kontrollü robotun görevleri yerine getirme yeteneğidir	İnsan, Robot, Görev, Yetenek

Akademik çalışma ve İş Dünyası tarafından açıklanan bazı YZ tanımlamaları aşağıda yer almaktadır.

Çizelge 2.1.2 Akademi ve İş Dünyası YZ Tanımları

Sıra No	Yazar	Yıl	İçerik	Yapay Zekâ Tanımı
1	JOHN MCCARTHY	2004	What Is Artificial Intelligence?	Akıllı makineler yapma bilimi ve mühendisliğidir [25].
2	HARUN PİRİM	2006	Yapay Zekâ	"İnsan aklının nasıl çalıştığını göstermeye çalışan bir kuram" [26].
3	HARUN PİRİM	2006	Yapay Zekâ	"Düşünme, anlama, faaliyete geçirmeyi sağlayacak bilgiişleme çalışmasıdır" [26]
4	YAVUZ KÖROGLU	2017	Yapay Zekâ'nın Teorik ve Pratik Sınırları	Yapay Zekâ, eldeki sorunun tanımı bilinir, fakat çözümün yöntemi (algoritması) bilinmezken, doğru ve verimli bir çözüm yöntemini çıkarım sayan, öğrenen, ya da keşfeden, insan eliyle üretilmiş sistemlerin tümüne verilen isimdir. Kısaca Yapay Zekâ, algoritma üretebilen otomatik sistemlerdir [27].
5	ESREF ADALI	2017	İtü Vakfı Dergisi Sayı 75 İnsanlaşan Makinalar-Yapay Zekâ	İnsan gibi düşünen ve davranan sistemler: Yapay Zekâ terimini öne-renlerin beklentileri, insan gibi düşü-nen ve dolayısıyla insan gibi davranan bilgisayarların geliştirilmesidir [28].
6	COSKUN SONMEZ	2018	Yapay Zekâ İçerikleri	YZ'nin amacı, normal olarak insan Zekâsını gerektiren görevleri yapabilecek makinalar yapmaktır [29].

7	COSKUN SONMEZ	2018	Yapay Zekâ İçerikleri	Yapay Zekâ araştırmalarının amacı, insan varlığında gözlemlediğimiz ve "akıllı davranış" olarak adlandırdığımız davranışları gösterebilen bilgisayarlar yapmaktır [29].
8	IBM	2020	Yapay Zekâ	Yapay Zekâ, insan zihninin problem çözme ve karar verme yeteneklerini taklit etmek için bilgisayarlardan ve makinelerden yararlanır [30].
.9	STUART J. RUSSELL, PETER NORVIG	2021	Artificial Intelligence: A Modern Approach	Algılama, akıl yürütme ve harekete geçmeyi sağlayan hesaplamalar [31].
10	ORACLE	2022	AI Nedir?	Görevleri yerine getirmek için insan Zekâsını taklit eden ve topladıkları bilgilere göre yinelemeli olarak kendilerini iyileştirebilen sistemler veya makineler anlamına gelir [32].
11	SAS	2022	Yapay Zekâ Nedir ve Neden Önemlidir	Yapay Zekâ (AI), makinelerin deneyimden öğrenmesini, yeni girdilere uyum sağlamasını ve insan benzeri görevleri gerçekleştirmesini mümkün kılar [33].

Çizelge 2.1.2 Akademi ve İş Dünyası YZ Tanımları

Otorite ve Akademi tarafından yapılan YZ tanımları direkt ve dolaylı olarak incelendiğinde; insan tarafından uzman yargısıyla geliştirilen sistemlerin, günlük yaşam içindeki süreçlere katkı sağlama yeteneğini, YZ olarak tanımlamak mümkündür.

Çizelge 2.1.3: YZ Tanımlarından Çıkarım

Tanımlanan Terim	Mevcut Tanımın Özeti	Mevcut Tanım için Anahtar Kelimeler
Yapay Zekâ	Sistemlerin yaşam içindeki süreçlere katkı sağlama yeteneği	İnsan, Süreç, Sistem, Sonuç, Yetenek

2.2 Makine Öğrenmesi

Öğrenme dürtüsü insanın gözlem yeteneğinin bir parçasıdır. İnsanlar günlük yaşamında bulunduğu ortam, toplum, sosyoekonomik çevre ve kültüre göre farklı öğrenme kuramlarını isteyerek ya da istemeyerek geliştirmektedir. Bu durum bize öğrenmenin bir süreç ve yolculuk olduğunu göstermektedir. Öğrenme kavramı genel olarak mekanik bir süreç olarak özetlenebilirken, insan öğrenmesini sahip olduğu duyu yeteneklerinden dolayı dinamik bir süreç olarak hayal edebiliriz. Aynı süreci makine öğrenmesi özeline indirecek olursak sahip olamadıkları duyu yetkinliklerinden dolayı statik bir sürecin parçası olmaktadır. Öğrenme kuramı ve insan öğrenmesi üzerine yapılan araştırmalar öğrenmenin uyarılma sonucu meydana getirdiği bir tepki olarak düşünmektedir.

Benzer şekilde aksiyon kabiliyeti olmayan nesnelerin uyarılması Newton'un üçüncü yasası olan etki tepki kuvvetinin ortaya çıkmasına sebep olmaktadır. Bu durum beraberinde nesnelerin tekrar yetisinin oluşmasına sebep olabilmektedir. Duran bir topa vurulması topa aktarılan enerjinin harcana kadar topun yuvarlanma hareketinin tekrar etmesine sebep olacaktır. Topun tek basına farklı yansımalara neden olmaması öğrenme yetisinin insanlara özgü bir davranış olmasını bizlere gösteren bir örnek olabilir. Bu durum insanlar tarafından farklı nesnelere makinelere aktarılabilir yüklenebilir.

Böylece MÖ terminoloji olarak makineden dolayı mekanik, öğrenmeden dolayı insan duygularını çağrıştıran kelimelerden türeyen içinde hem teknik hem de öğrenme dürtüsü olan sistemlerin birleşimi olarak özetlenebilir. Makine öğrenmesi temel olarak YZ uygulamalarının alt kümesi olarak görülüp özetlenebilirken YZ kavramının dışında konumlandırmakta mümkündür. Önde gelen bazı teknoloji şirketleri tarafından makine öğrenmesi için yapılan tanımlamalar çizelge 2.2.1 üzerinde gösterilmektedir.

Çizelge 2.2.1 Akademi ve İş Dünyasının Tanımları

Kurum	Tanım
IBM	Makine öğrenmesi, insanların öğrenme şekillerini taklit etmek için veri ve algoritmaların kullanımına odaklanıp doğruluğunu kademeli olarak artıran bir Yapay Zekâ (AI) ve bilgisayar bilimi dalıdır [34].
Microsoft	Makine öğrenmesi (ML), bir bilgisayarın doğrudan yönergeler olmadan öğrenmesine yardımcı olmak için matematiksel modelleri kullanma işlemidir [35].
Amazon	Makine öğrenimi, bilgisayar sistemlerinin açık talimatlar yerine düzenlere ve çıkarıma bağlı olarak görevleri gerçekleştirmek için kullanacağı algoritmalar ve istatistiksel modeller geliştirme bilimidir [36].
SAS	Makine öğrenimi, analitik model oluşturmayı otomatikleştiren bir veri analizi yöntemidir. Sistemlerin verilerden öğrenebileceği, örüntüleri tanımlayabileceği ve minimum insan müdahalesi ile kararlar alabileceği fikrine dayanan bir Yapay Zekâ dalıdır [37].
Oracle	Makine öğrenimi (ML), tükettikleri verilere göre öğrenen ya da performansı iyileştiren sistemler oluşturmaya odaklanan bir Yapay Zekâ (AI) alt kümesidir [38].

2.3 Algoritma

YZ ve MÖ terimlerinden bahsedildiğinde akıllara sonradan anlamlandırması görece daha karmaşık olan algoritma kavramı gelmektedir. Algoritma kavramının kullanımı teorik olarak çok eskilere dayanmaktadır. TDK göre algoritma, "orta çağda ondalık sayı sistemine göre, son zamanlarda ise iyi tanımlanmış kuralların ve işlemlerin adım adım uygulanmasıyla bir sorunun giderilmesi veya sonuca en hızlı biçimde ulaşılması işlemidir". Tanımdan anlaşılacağı gibi algoritmalar birer süreçtir ve makine öğrenmesi disiplini altında öğrenme yetisi olarak kullanılmaktadır.

Microsoft algoritmaları, "kişilerin, karmaşık veri kümelerini keşfetmesi, analiz etmesi ve bunlarda anlam bulmasına yardımcı olan kod parçacıkları ve Her algoritma, bir makinenin belirli bir hedefi gerçekleştirmek için izleyebileceği sınırlı ve belirli, adım adım ilerleyen yönerge kümesi" olarak tanımlamaktadır. Oracle algoritmaları "makine öğrenimine güç sağlayan motorlar" olarak tanımlamaktadır. İlgili tanımlamalardan anlaşılacağı gibi algoritmalar olmadan makine öğrenmesinin pek mümkün olmayacağı ve makine öğrenmesinin belirli adımları yönergeler içinde takip eden bir süreç olduğunu düşünebiliriz. Bu durum makine öğrenmesinin gücünü ve kaderini algoritmalardan aldığını göstermektedir.

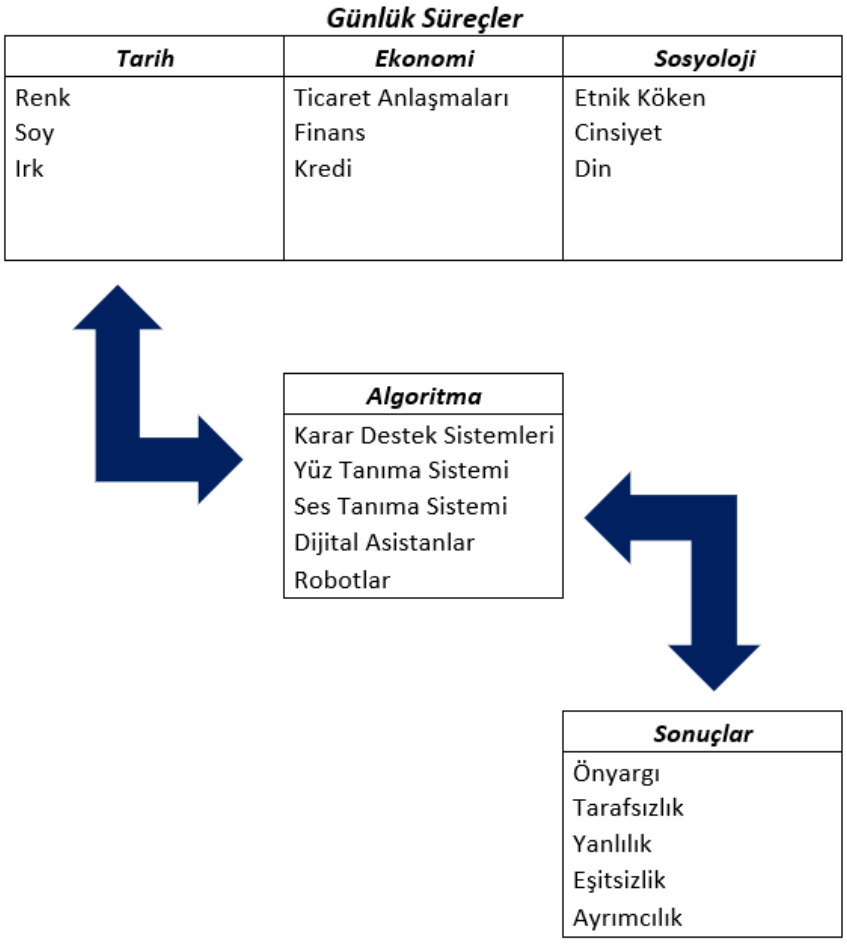
2.4 Algoritmik Önyargı

Albert Einstein'ın söylediği gibi "aynı şeyleri tekrar tekrar yapıp farklı sonuçlar beklemek çok olası değildir". Algoritmanın oluşturulma ve çalışma prensipleri göz önüne alındığında farklı aşamaların tekrar tekrar çalışıp aynı sonuçları veya çıktıları üretmesi beklenmektedir. Algoritmaların tasarımı, etki alanı, süreci ve yörüngesi kaynaklı olabileceği gibi etki alanı dışına çıkamaması mevcut durumu farklılaştıramamasından dolayı yanlamasına neden olabilecektir. Bu durum beraberinde algoritma önyargısı (Algorithmic Bias) kavramını ortaya çıkarmıştır.

Önyargı problemi günlük hayatımızın her alanında olmakla birlikte farklı disiplinler hatta bilimlerin de problemi olmaktadır. Siyasi, Ekonomi, Sosyoloji ve Tarih gibi dalların çoğunda önyargı problemleri karşımıza çıkmaktadır.

Mevcut süreçlerdeki problemin algoritmalar tarafından karar destek, yüz tanıma, robotik ve sistemlere aktarılması muhtemeldir.

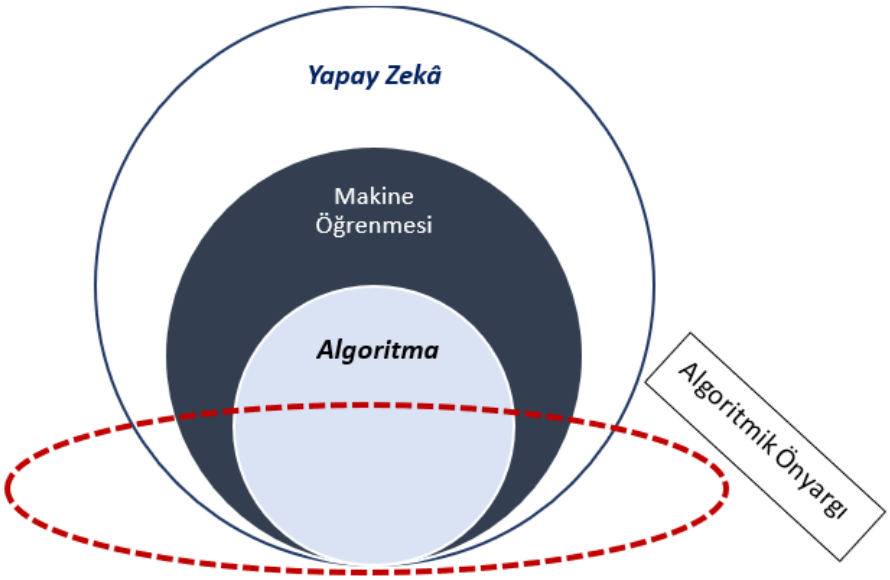
Algoritmaların aynı süreci tekrar etmesiyle günlük süreçlerdeki önyargı, eşitsizlik vb. durumların artırması ve şekil 2.4.1'de görüldüğü gibi çift yönlü kısır döngüye girmesi muhtemeldir.



Şekil 2.4.1: Günlük Süreçlerde Algoritma Döngüsü

Şekil 2.4.1'de görüldüğü gibi insanların günlük yaşamının bir parçası olan problemleri farkında olarak ya da olmadan algoritmalara taşındığını ve problemleri günlük hayatının bir parçası olarak destek sistemleri ile devam ettirdiğini görebiliriz.

Algoritmik önyargı kavramını YZ, MÖ ve Algoritma arasında şekil 2.4.2'deki gibi konumlandırabiliriz.



Şekil 2.4.2: YZ, MÖ, Algoritma ve Algoritmik Önyargı Konumlanması

Algoritmik Önyargı kavramına günlük hayatımızda kullandığımız at gözlüğü ile bakmak deyimini örnek gösterebiliriz. Etrafımızda olan olayları tamamen anlayamamamız olaylara karşı taraflı yaklaşmamıza neden olabilir. Bazen farkında olmadan insanın sahip olduğu kültür ve düşünce yapısı iyi niyet çerçevesinde bile insanların kutuplaşmasına neden olabilir. İnsan ufkunun bile kısıtlı kaldığı bu durumlarda makinelerin öğretilen algoritmaların dışına çıkamaması aslında şaşırılması gereken değil beklenen bir durumdur. İnsanların kendilerini anlatabildiği kadar karşı tarafta etki bırakabileceğini düşünecek olursak insanlar tarafından geliştirilen algoritmaların yine insanların kendi ufkunu aktardığı kadar çalışması beklenmektedir. Algoritmaların tekrar tekrar çalışıp öğrenme yetisi oluşturmasını çığ düşmesi olayına benzetebiliriz. Kar tanesinin birikerek kayması sonucu meydana getirdiği kar topu kütlesi tekrar ettikçe büyümekte ve önüne çıkan nesnelere zarar vermektedir. Algoritma önyargısı ürettiği sonuçlar ile çığ etkisi yaratabilir ve belirli kesimlerin zarar görmesi veya toplumsal eşitsizliğin oluşmasına neden olabilir.

Farklı sektörlerden popüler şirketlerin karar destek hizmetlerindeki algoritmik önyargılar aşağıda Çizelge 2.4.1 içinde ürün bazlı örneklendirilmiştir.

Çizelge 2.4.1 Algoritmik Önyargı Örnekleri

Ürün	Kullanım Amacı	Olay	Algoritmik Önyargı
Microsoft – Tay	YZ Robotu	Kullanıcı Sohbeti	Küfür, Cinsiyet, Irkçılık
Amazon – Alexa	YZ Robotu	Kullanıcı Sohbeti	Cinsiyetçilik, Ayrımcılık
Tesla - Model 3	Otonom Süreç	Tanımsız Cisim Kazası	Kavram Kargaşası
Facebook - AI	Yüz Tanıma	Siyahileri farklılaştırma	Irk Ayrımı
Google - Translate	Doğal Dil İşleme	Diller arası gelişmişlik seviyesi	Cinsiyetçilik, Ayrımcılık

Çizelge 2.4.1 içinde dar kapsamlı süreçleri kusursuz tekrar etmesiyle popülerlik kazanmış günümüzde yenilikçi, inovasyon veya çığır açıcı olarak adlandırdığımız çeşitli ürünlerin kapsam ya da kullanım alanın genişlemesiyle algoritmik sıkıntıları yaşadığı görülmüştür. Tekrar edilen süreçlerin farklılaşması algoritmaların adaletsizlik veya eşitsizlik yaratmasına sebep olmuştur. Bu durum beraberinde toplumsal ayrışmalara neden olmuştur. Bu ürünler dışında benzer şekilde kullanıcılar arası eşitsizlik yaratan yaşanmış olaylara aşağıda örnekler verilmiştir.

Amerika Birleşik Devletleri'nde (ABD) kullanılan bazı suç tahmini algoritmaları, yoksul kişileri daha fazla suçlu olarak göstermektedir. Northpointe tarafından geliştirilen COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algoritması, yoksul kişilerin yeniden suç işleme olasılıklarını daha yüksek seviyede tahmin etmektedir [39].

Amazon, işe alım süreci için hazırladığı Yapay Zekâ destekli mülakat yazılımın kadın adaylara karşı önyargılı bir yaklaşım sergilediği ortaya çıkmıştır. Yazılım, kadınların ifadelerini, erkeklere göre daha az değerli görmüştür [40].

ABD'de 2018 yılında kullanılan yüz tanıma sistemleri, siyahileri birey olarak tanımlama olasılığının diğer ırklara göre daha düşük doğruluk oranına sahip olduğu ölçülmüştür [41].

Amazon 2016 yılında, kullanıcılara göre fiyat farklılığı gösteren pazarlama algoritması kullanmıştır. Sunulan algoritmaların erkeklerin sıklıkla ziyaret ettiği ürünleri, kadınlara kıyasla daha yüksek seviyeden satışa sunulduğu tespit edilmiştir [42].

2019 yılında gerçekleştirilen bir araştırmaya göre Facebook'un reklam algoritmasının, iş ilanlarını kadınlara kıyasla erkeklere daha fazla gösterdiği tespit etmiştir [43].

2019 yılında yapılan bir başka araştırmaya göre IBM'in Doktor Watson sağlık hizmetleri algoritması, beyaz hastalara, siyahi hastalara mukayese daha doğru teşhis koyduğu tespit edilmiştir [44].

Uber'in fiyatlandırma algoritmasının, gelir ortalaması yüksek olan bölgelerde aynı mesafedeki diğer yollara göre daha fazla fiyat tarifesi uyguladığı tespit edilmiştir [45].

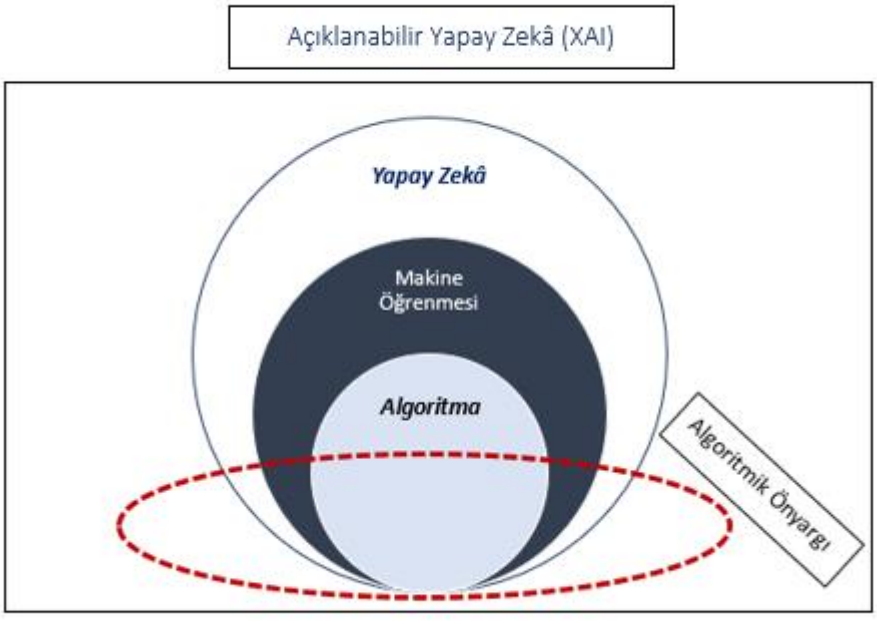
Yukarıda bahsedilen örnekler, algoritma yanlılığı veya algoritmik önyargı olarak tanımladığımız sorunu farklı alan ve sektörler için ortaya koyduğu eşitsizliği göstermektedir. Algoritmalar, temelde veri girdilerine dayanarak otonom kararlar verirler ve girdiler insan kaynaklı olduğu için, farkında olmadan algoritmaların önyargı veya yanlılık oluşturmasına sebebiyet verebilir.

Verilen örnekler incelendiğinde, algoritmaların sınıflandırma yaparak; bir gruba karşı avantaj sağlarken, diğer gruba (yoksul kişiler, kadınlar, siyahiler vb. gibi) ayrımcılık yaptığı görülmektedir. Algoritmaların kararlarının etik, adil ve doğru olduğundan emin olmak için algoritmaların geliştirme sürecinde çok daha fazla şeffaflık, açıklık ve insan denetimi gereklidir. Bu nedenle günümüzde "Açıklanabilir Yapay Zekâ (XAI)" yaklaşımı ortaya atılmış ve yavaş yavaş literatür içinde kavramsal olarak yer edinmeye başlamıştır.

2.5 Açıklanabilir Yapay Zekâ (XAI)

Açıklanabilir Yapay Zekâ (XAI), algoritmaların karar verme süreçlerini daha şeffaf ve anlaşılır hale getirmek için önerilen yeni bir yaklaşımdır [84]. XAI, algoritmaların kararlarına etki eden faktörleri, verileri ve işlemleri anlaşılır bir şekilde açıklamayı ve bu sayede kararların doğruluğunu, adaletini ve güvenilirliğini artırmayı hedefler. Yukarıda bahsedilen "algoritmik önyargı" probleminin önlenmesinde XAI yaklaşımı çözüm odaklı önem taşıyabilir. XAI sayesinde algoritmaların nasıl ve neden karar verdiği anlaşılır hale gelir ve bu sayede önyargılı kararların olması halinde uzmanlar tarafından daha kolay tespit edilebilir. Bu sayede kavranması zor bir iç işleyişi olan algoritmalar (black box algorithms) kendilerini şeffaf ve adil algoritmalara bırakır. XAI ayrıca, kullanıcıların algoritmalara daha fazla güven duyulmasını sağlamayı amaçlamaktadır. Kullanıcılar algoritmaların nasıl çalıştığı hakkında fikir sahibi oldukça, algoritma kararlarına daha fazla güven duyabilirler ve bu sayede algoritmaların kabul edilmesi ve kullanımı artabilir.

XAI, şekil 2.4.2'de tanımlanan algoritmik önyargının konumlamasının üstüne şekil 2.5.1 gibi çerçeveleme görevi görerek Yapay Zekâ teknolojilerinin etik ve adaletli bir şekilde kullanılmasına olanak tanımayı amaçlar.



Şekil 2.5.1: XAI Konumlandırılması

3. LİTERATÜR TARAMASI

MÖ ve YZ çalışmaları kapsamında günümüzde önemli çalışmalar bulunmaktadır. Çalışmaların artmasıyla beraber gelişen literatür çözümler getirdiği beraberinde MÖ ve YZ önyargısı gibi problemlerin oluşmasına neden olabilmektedir.

MÖ ve YZ konularında önyargı kavramı ve çalışmaları yakın zamanda ilgi odağı olmuştur [46] (Pessach ve Shmueli, 2022).

Köken, etnik ve benzeri hassas verilerin girdi olduğu yapa Zekâ modellerinde karar mekanizmalarında önyargının oluşabileceği düşünülmektedir [47] (Barocas, Hardt ve Narayanan, 2017).

MÖ ve YZ uygulamalarının kullanımının artmasıyla insanlara karşı önyargı gibi beklenmedik sosyal sonuçlar artabilir [48] (Caton ve Haas, 2020).

MÖ ve YZ belirli demografik gruplar için önyargılı/haksız kararlar verebilir ve eşitsizliğe sebep olabilir [49] (Zhao ve diğerleri, 2022).

Mevcut MÖ ve YZ süreçlerindeki önyargı gelecekteki olaylara miras kalabilir ve haksız kararların artmasına yol açabilir [50] (Obermeyer ve diğerleri, 2019).

MÖ ve YZ çalışmalarında popülasyonu temsil etmeyen veya edemeyen örneklem kümeleri karar mekanizması süreçlerinde önyargıya sebep olabilir [51] (Plumed ve diğerleri, 2019).

MÖ ve YZ çalışmalarındaki kaynaklar hakkındaki bilgi eksikliği ve mevcut bilgilerin dağınıklığı önyargıya sebep olabilmektedir [52] (Fabris ve diğerleri, 2022).

Açıklanabilir tekniklerler ile algoritmaları insanların anlayabileceği seviyeye gelmesini ve kara kutu olarak görünen MÖ ve YZ süreçlerinin aydınlatılması hedeflenmiştir [53] (Ras ve diğerleri, 2022).

MÖ ve YZ alanında eşitsizlikle ilgili sıklıkla kullanılan veri kümeleri üzerinde algoritmik eşitsizlik uygulanmıştır [54] (Pessach ve Shmueli, 2020).

Denetimli öğrenme sistemlerinde veri toplama ve etiketleme tekniklerinin algoritmik önyargı üzerine etkisi ölçülmesi hedeflenmiştir [55] (Li ve diğerleri, 2022).

Basit aktif örnekleme ve yeniden ağırlıklandırma stratejileri ile algoritmik eşitsizliğin optimize edilmesi hedeflenmiştir [56] (Abernethy ve diğerleri, 2022).

Çizelge 3.1: Sistematik Literatür Tarama Özeti

Tarih	Yazarlar	Anahtar Kelimeler	Araştırmanın Konusu	Kullanılan Modeller / Algoritmalar	Analizler	Amaç / Hedef / Katkı
2020	Caton, Haas	Tarafsızlık, Hesap Verebilirlik, Şeffaflık, Makine Öğrenmesi	Makine öğrenmesinde önyargı	Gözetimsiz makine öğrenmesi ve doğal dil işleme	Eşitlik Metrikleri / Endeksleri	Makine öğrenmesi önyargılarının hafifletilmesi ve tarafsızlık sağlamaya yönelik yaklaşımlar sunmayı amaçlamaktadır [57].
2020	Biswas, Rajan	Tarafsızlık, Makine Öğrenmesi, Modeller	Makine öğrenmesi eşitsizliği üzerine sistematik araştırma	Gözetimli, Gözetimsiz ve Destekli Makine Öğrenmesi	Eşitlik Metrikleri	Makine öğrenmesi tarafsızlığına yönelik katkıların ön plana çıkmasını amaçlamıştır [58].

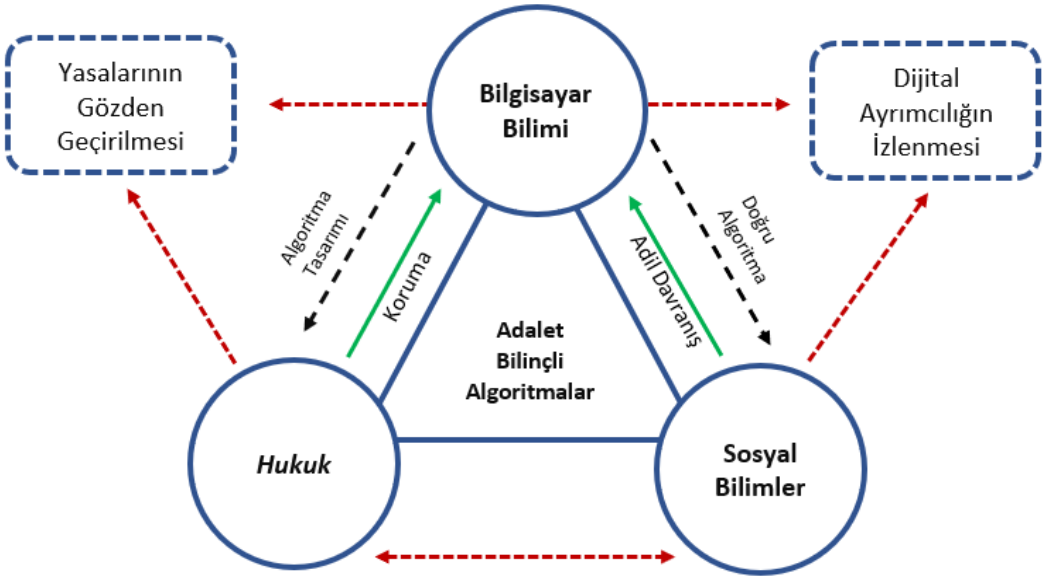
2021	Quy ve diğerleri	Karşılaştırmalı Veri Kümeleri, Önyargı, Ayrımcılık, Tarafsızlık farkındalığına sahip makine öğrenimi	Makine öğrenmesinde önyargının farkındalığı	Bayes Ağı	Açıklayıcı Faktör Analizleri	Veri kümeleri analizlerindeki makine öğrenmesi önyargılarının anlaşılması ve gösterilmesi amaçlanmıştır [59].
2021	Fuster ve diğerleri	Makine Öğrenmesi, Ev Kredisi, Farklı Etki, Irk	Makine öğrenmesinin kredi piyasalarındaki önyargısı	Logit Modeller, Ağaç bazlı algoritmalar	Temerrüt Olasılığı Analizi	Farklı etnik grupların (ırk) kredi piyasasında karşılaşacağı eşitsizliği göstermeyi amaçlamıştır [60].
2021	Kasmi	Kredi Puanlama, Adil makine öğrenimi, Algoritmik tarafsızlık, Ayrımcılık	Makine öğrenmesinin kredi skorlama sürecindeki önyargısı	Gözetimli makine öğrenmesi	Kredi Skorlama Analizleri	Tarafsızlık kriterlerinin kredi puanlamalarındaki yeterliliklerini incelemeyi ve tarafsızlığı sağlayacak algoritmik seçenekleri göstermeyi amaçlamıştır [61].

2021	Szepannek	Puanlama, Makine Öğrenmesi, Nedensel çıkarım, Alman kredi verileri, Algoritmik Tarafsızlık	Makine öğrenmesinin risk skorlamasındaki tarafsızlığının zorluğu	Gözetimli makine öğrenmesi	Nedensel Çıkarım ve İstatistiksel eşitlik Oranları	Simülasyon yardımıyla, makine öğrenmesi tarafsızlığını ve tahmin doğruluğunu iyileştirmek hedeflenmiştir [62].
2022	Makhlouf ve diğerleri	Tarafsızlık, Makine öğrenmesi, Nedensellik, Nedensel çıkarım, Müdahale, Karşıolgu	Makine öğrenmesinde önyargı	Yapısal nedensellik modelleri	Nedensellik, İstatistiksel eşitlik oranları	Uygun tarafsızlık kavramının seçilmesine yardımcı olmak amaçlanmıştır [63].
2022	Cimatec,	Önyargı, Tarafsızlık, Makine Öğrenmesi, Yapay Zekâ	Makine öğrenmesindeki önyargılar	Gözetimli, Gözetimsiz ve Destekli Makine Öğrenmesi	Bibliyometrik analizi ve Tekrarlanan En Küçük Kareler Yöntemi (RSL)	Makine öğrenmesinde önyargı ve adaletsizlik kavramlarını ön plana çıkarmayı amaçlamaktadır [64].

2022	Choudhary ve diğerleri	Grafik, Sosyal ağlar, Düğüm yerleştirme, Önyargı, Adalet,	Makine öğrenmesindeki önyargıların görselleştirilmesi	Gözetimli, Gözetimsiz ve Destekli Makine Öğrenmesi	Ağ Analizleri	Grafik madenciliği teknikleri ile model önyargı eğilimlerinin incelenmesi amaçlanmıştır [65].
2022	Dablain ve diğerleri	Tarafsızlık, Dengesiz Veri, Makine Öğrenmesi	Makine öğrenmesindeki önyargılar ve çözümler	Gözetimli Makine Öğrenmesi	Yüksek sayıda örnekleme (Oversampling)	Makine öğrenmesi önyargısı ve tarafsızlığı arasındaki ilişkilerin gösterilmesi hedeflenmiştir [66].
2022	Fabris ve diğerleri	Algoritmik Tarafsızlık, Veri Çalışmaları	Makine Öğrenmesindeki önyargı ve veri	Makine Öğrenmesi Teknikleri	Karşılaştırmalı Analiz	Algoritmik tarafsızlık araştırmalarında kullanılan veri setlerinin incelenmesi amaçlanmıştır [67].
2022	Raimondi ve diğerleri	Eşitlik, Algoritma, Makine Öğrenmesi	Algoritmik tarafsızlık için çözüm önerileri	Lojistic Regresyon Modeli	İstatiksel Anomiler	Algoritmaların tahmin gücünün artırılması ve eşitsizliğin giderilmesi amaçlanmıştır [68].

4. MAKİNE ÖĞRENMESİ DESTEKLİ KARAR DESTEK SİSTEMLERİ ÜZERİNE KREDİ DEĞERLENDİRME

Algoritmik önyargı veya başka bir söylemle adil olmayan algoritmalar sadece Yapay Zekâ disiplini değil farklı disiplinlerin hatta bilimlerin problemi olabilmektedir. Zliobaite I. tarafından 2017 yılında gösterilmeye çalışılan disiplinler arası algoritma eşitsizliği etkileşimi aşağıda özetlenmeye çalışılmıştır [70]. Algoritmik önyargının disiplinler arası etkileşimi eşitsizliklerin ölçümü ve tanısı için önemli olabilmektedir. Algoritmaların tarafsızlık gösterebilmesi için doğru kullanım, tasarım, uygun davranış gibi konuların doğru saptanması gerekmektedir.

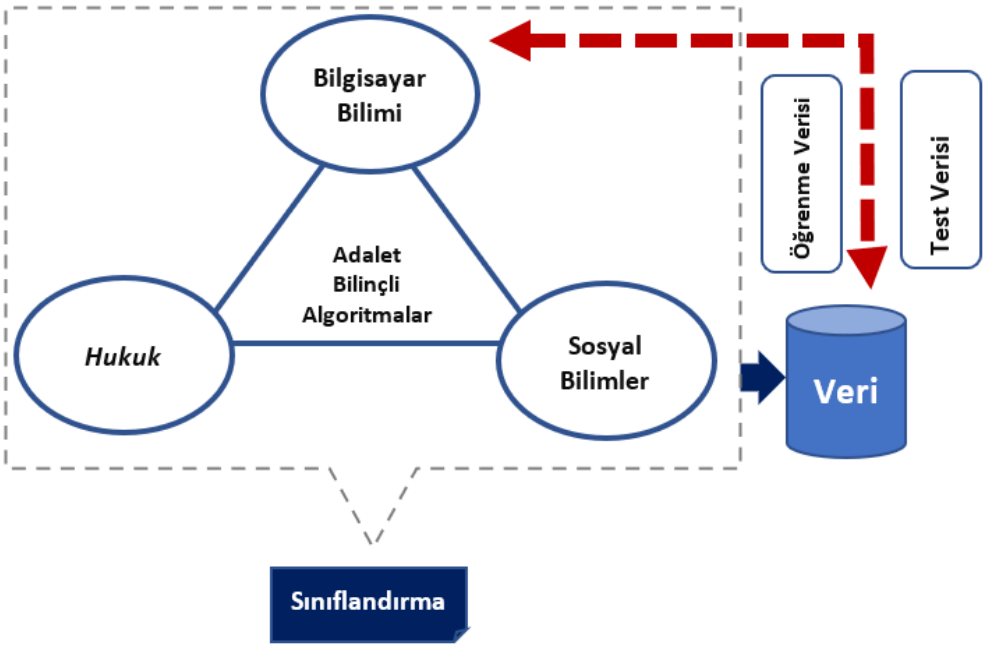


Şekil 4.1: Disiplinler Arası Algoritma Köprüsü

Makine öğrenmesi süreçlerine temelde istatistiksel ayrım ve/veya sınıflandırma ölçümü olarak bakılabilir. İstatistiksel ayrım mevcut veriler dahilinde benzerlik, farklılık gibi sınıflandırmaların tespit edilmesi olarak özetlenebilir. İstatistiksel çıkarımların başarısı için analiz edilen veri kümesinin önemi büyüktür. Veri kümesinin çok yüksek veya düşük seviyelerde örneklem içermesi hatalı çıkarımlara sebep olabilir.

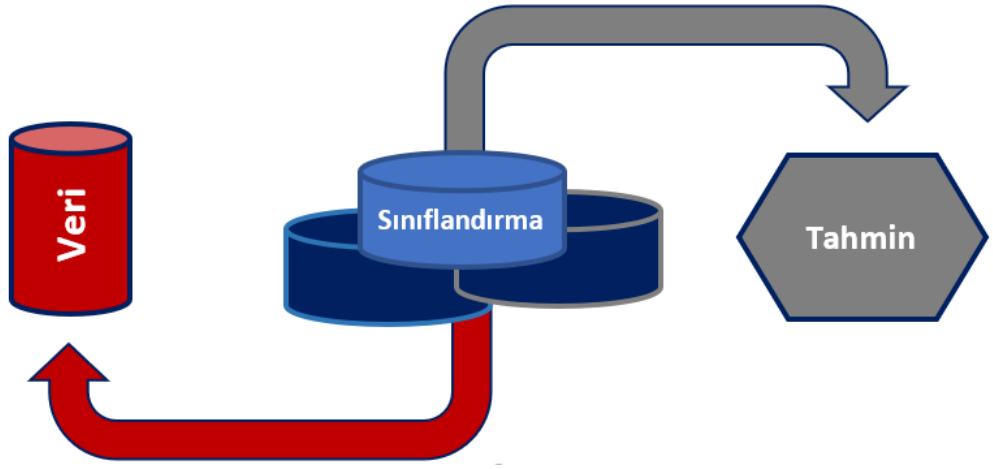
Veri kümesinin adil olmayan etiketler içermesi yine hatalı çıkarımları tetikleyebilir. Makine öğrenmesine uygun olmayan veriler algoritmaların önyargı yaratmasına sebep olacaktır.

Bu kapsamda yukarıda gösterilen şekil 4.1'i basitleştirip ve verinin önemi ile genişletilince baştan sona doğru tüm süreci şekil 4.2'deki gibi sınıflandırma olarak göstermemiz mümkündür.



Şekil 4.2: Disiplinler Arası Algoritma Etkileşime Yeni Bakış

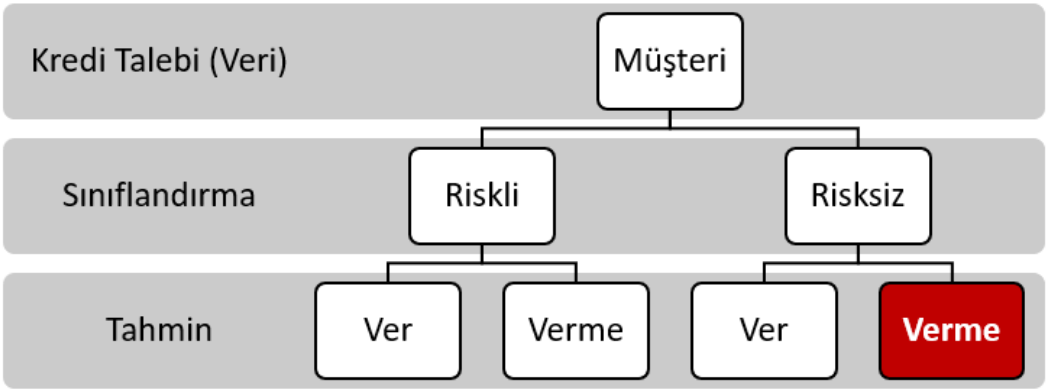
Şekil 4.1'den çıkarım yapılacağı üzere algoritmik önyargının aslında sınıflandırma bir çeşit sınıflandırma problemi olduğu görülmektedir. Sınıflandırma problemlerinin saptanmasında tek bir yöntem bulunmamakla birlikte literatür taramasında gördüğümüz gibi farklı ölçüm metrikleri ve istatiksel yöntemler bulunmaktadır. Adalet tanımının birden fazla olması birden fazla ölçüm yöntemi olmasını destekler niteliktedir. Sınıflandırmanın yanlaması hatalı tahmin ve eşitsizlik problemlerine yol açacaktır. Şekil 4.2'yi sadeleştirip bir adım daha devam ettirecek olursak bütün süreci baştan sona özetleyen akışı aşağıda şekil 4.3'te görebiliriz.



Şekil 4.3: Algoritma Döngüsü

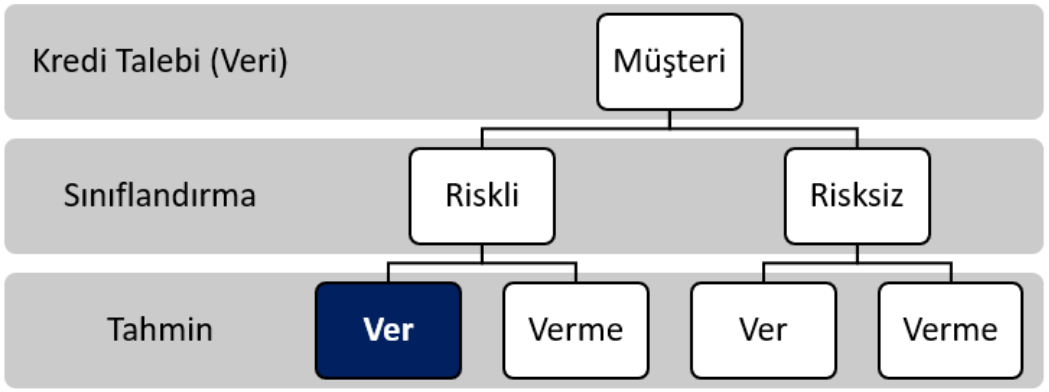
Şekil 4.3'te görüldüğü gibi veri kümesi seçimi algoritmanın tahmin gücü için önemli bir seçimdir. Bu çalışmada veri kümesi belirlenmesinde halka açık veri kümesi olan Alman kredi veri seti (AKVS) kullanılacaktır. AKVS kullanımın tercih edilmesinin sebepleri sırasıyla; halka açık olması, referans sayısının fazla olması, makine öğrenmesi çalışmalarında yaygın olması, veri setinin detaylı olması, değişken sayısının fazla olması, sınıflandırma probleminde kullanılması ve finans sektöründe olmasından dolayı karar destek mekanizmasında önemli etkisinin olmasıdır.

AKVS bir banka tarafından kredi başvurusu sürecinde, başvuru sahibinin özelliklerine göre kredi kullanımı izin verilip verilmeyeceğini açıklamak için kullanılmaktadır. Tahmin sonucuna göre banka müşterinin riskli olduğunu düşünmekteyse krediyi onaylamaması gerekmektedir. Bankanın müşteriye kredi vermemesi kararı beraberinde fırsat eşitsizliği, adaletsizlik ve önyargı gibi konuları gündeme getireceği gibi Bankanın müşteri riskini doğru tespit edememesi ve kredi vermesi yine aynı konulara sebep olabileceği gibi bankanın finansal zarara uğramasına da neden olabilir. Kredi kararının etkileri müşteri ve banka tarafından ağaç dallanması ile aşağıda şekil 4.4 ve 4.5'te özetlenmiştir.



Şekil 4.4: Algoritma Karar Ağacı (Risksiz-Verme)

KDS sonucunda kredi riski düşük müşteriye kredi verilmemesi ilgili müşteri için fırsat eşitsizliğine sebep olurken, banka içinde gelir kaybına neden olmaktadır.



Şekil 4.5: Algoritma Karar Ağacı (Riskli-Ver)

KDS sonucunda kredi riski yüksek müşteriye kredi verilmesi ilgili müşteri için pozitif ayrımcılığa sebep olurken banka içinde yüksek kredi riski ile karşı karşıya olmasına neden olmaktadır. Ağaç dallanması şekillerinde görüldüğü üzere KDS süreçlerinde algoritmik yanlamanın olması olasıdır. Bu kapsamda algoritmaların yanladığını düşüncesine açıklık getirilmesi adına aşağıdaki araştırma sorularının test edilmesi hedeflenmiştir.

Kredi karar destek sistemlerinde algoritmik eşitsizlik yoktur.

Kredi karar destek sistemlerinde algoritmik eşitsizlik vardır.

Araştırma sorularının mevcut veri seti üzerinde istatistiksel analizler ve eşitsizlik metrikleri ile R ve Python programları kullanılarak sayısallaştırılıp ölçülmesi hedeflenmektedir.

4.1 Veri Toplama

Tez kapsamında kredi karar süreçlerinde kullanılan makine öğrenmesi algoritmaların önyargı veya eşitsizliğe neden olup olmadığı halka açık olan Alman Kredi Veri Seti (AVKS) (German Credit Data) üzerinde test edilmesi planlanmıştır. AVKS'nin tercih edilmesinin ana nedeni bireylerin finansal ve kişisel verilerinin toplamasının zor ve denetimli olduğu bir ortamda araştırmacılara açık kaynaklı bir seti olmasıdır. Bu nedenle makine öğrenimi ve veri madenciliği alanında sıklıkla kullanılmış "standart" bir veri niteliğindedir.

Alman Kredi Veri Seti (German Credit Data), 1989 yılında Prof. Dr. Hans Hofmann tarafından oluşturulmuştur. Bu veri seti, bankacılık sektöründe kredi risk analizi ve kredi verme kararlarının otomatikleştirilmesi gibi konularda kullanılmak üzere tasarlanmıştır. Hazırlanan bu veri seti daha sonra farklı disiplin ve alanlarda analiz çalışmalarında kullanılmıştır. AVKS kullanılarak hazırlanan çalışmalar Scopus, Elsiver, IEEE Xplore gibi saygın platformlarda yer almaktadır. Aynı zamanda AVKS verileri ile hazırlanmış çalışmalar farklı disiplinlerde alanında önde olan dergiler (Journal of Intelligent Information Systems, Journal of Business Research, Journal of Risk Research, Journal of Retailing, Journal of Banking and Finance, Journal of Credit Risk) tarafından kabul edilmiş ve yayınlanmıştır. AKVS, açık kaynaklı bir veri seti olduğu için araştırmacılar ve veri bilimciler tarafından kolayca ve sıklıkla kullanılmaktadır.

AKVS, anonim bir bankanın müşterileriyle ilgili finansal bilgileri içeren gerçek veri setidir. Bu veri seti, banka müşterilerinin kredi taleplerini karar vermeye yardımcı olmak için toplanmıştır. Veri seti, 1000 müşteriye ait 20 farklı özellik içermektedir.

Bu özellikler arasında müşterinin yaşı, cinsiyeti, medeni durumu, geliri, mesleği, kredi tutarı, kredi süresi, kredi geçmişi, varlık durumu, yabancı ülkeye seyahat etme durumu ve benzeri gibi bilgiler yer almaktadır.

Özellikle, kredi değerlendirme modellerinin geliştirilmesi, risk analizi yapılması, müşteri segmentasyonu çalışmaları gibi konularda sıklıkla başvurulan bir veri setidir. AKVS hakkında genel bilgiler çizelge 4.1.1 ve 4.1.2'de özetlenmiştir.

Çizelge 4.1.1: AKVS Özellikleri

Alman Kredi Veri Setinin Özellikleri	
Küme Tipi:	Çok Değişkenli
Nitelik:	Kategorik ve Sayısal
Sektör:	Finans
Kullanım Alanı:	Sınıflandırma
Örnek Sayısı:	1000
Değişken Sayısı:	20

AKVS değişkenleri hakkında detay açıklamalar çizelge 4.1.2'de gösterilmiştir.

Çizelge 4.1.2: AVKS Açıklama

Değişkenler	Değişken Tipi	Açıklama
1	Kategorik	Mevcut Çek Ödemeleri
2	Sayısal	Kredinin Vadesi
3	Kategorik	Kredi Geçmişi
4	Kategorik	Kredi Kullanım Amacı
5	Sayısal	Kredi Tutarı
6	Kategorik	Mevduatı Var mı?

7	Kategorik	Çalışma Süresi
8	Sayısal	Taksit Tutarı / Mevcut Gelir
9	Kategorik	Cinsiyet
10	Kategorik	Başka Borcu Var mı?
11	Sayısal	Oturum Süresi
12	Kategorik	Mal Varlığı
13	Sayısal	Yaş
14	Kategorik	Kredi Dışında Borcu Var mı?
15	Kategorik	Oturum Bilgisi Kira / Mülk?
16	Sayısal	Mevcut Kredi Sayısı
17	Kategorik	Çalışma Durumu
18	Sayısal	Baktığı Kişi Sayısı
19	Kategorik	Telefon Faturası Var mı?
20	Kategorik	Yabancı İşçi mi?

Çizelge 4.1.2: AVKS Açıklama

AKVS değişken kümesi incelendiğinde geçmiş bölümlerde belirtilen ayrımcılık ve adaletsizliğe neden olabilecek Yaş, Cinsiyet, Irk gibi değişkenler direkt olarak dikkat çekmektedir. Başlangıç olarak Yaş ve Cinsiyet üzerinden verileri üzerinde yapılacak testler ile algoritmik eşitsizlik olup olmadığı test edilecektir.

4.2 Veri Analizi

Kredi veri setinde yaş ortalamasının altında kalanlar, ortalamanın üstünde kalanlara göre daha az tecrübeli olması ve cinsiyeti kadın olanların, erkek olanlara göre negatif ayrışması nedeniyle adaletsizliğe, eşitsizliğe ve algoritma önyargısına neden olabilmektedir. Tahmin sonuncunun kredi verilenler için pozitif (p), verilmeyenler için negatif (n) ve ayrıcalık olan küme X, olmayan kümenin Y olacağı varsayılırsa her değişken için 2 × 2 şeklinde matriks oluşmaktadır. İlgili matris örneklemi aşağıda Çizelge 4.2.1'de genel ve çizelge 4.2.2'de her değişken için ayrı ayrı detay matriks gösterilmiştir.

Çizelge 4.2.1: Genel Hata Matrisi

	Doğru	Yanlış
Pozitif	DP	YP
Negatif	DN	YN

Çizelge 4.2.2: Detay Hata Matrisi

	Ayrıcalıklı (X)	Ayrıcalıklı Olmayan (Y)
Kredi Verilen (p)	X(p)	Y(p)
Kredi Verilmeyen (n)	X(n)	Y(n)

$$Y(t) = Ayrıcalıklı\ Olmayan\ Tum\ Grup$$
$$Y(p) = Ayrıcalıklı\ Olmayan\ Kredi\ Verilen\ Grup$$
$$Y(n) = Ayrıcalıklı\ Olmayan\ Kredi\ Verilmeyen\ Grup$$
$$X(t) = Ayrıcalıklı\ Olan\ Tum\ Grup$$
$$X(p) = Ayrıcalıklı\ Olan\ Kredi\ Verilen\ Grup$$
$$X(n) = Ayrıcalıklı\ Olan\ Kredi\ Verilmeyen\ Grup$$

Yaş ve Cinsiyet verilerinin R ve Python programlarında IBM tarafından geliştirilen AIF360 kütüphane ve ilgili kütüphane ve literatür taraması sonucu yaygın olarak kullanılan 4 metrik için analiz edilmiştir.

İstatistiksel Parite Farkı (Statistical Parity Difference)

İstatistiksel Parite Farkı (Statistical Parity Difference), mevcut gruptaki bireylerin özellikleri (cinsiyet, yaş vb.) nedeniyle karşılaştığı ayrımcılığı ölçmek için kullanılan istatistiksel bir metriktir. İstatistiksel Parite Farkı, bir grubun maruz kaldığı olumlu veya olumsuz bir sonucun oranının, başka bir gruba kıyasla ne kadar farklı olduğunu gösterir. İstatistiksel Parite Farkı, cinsiyet, ırk, yaş, etnik köken gibi birçok faktör için kullanılabilir. Ancak, sadece iki grup karşılaştırmasında kullanılır.

İstatistiksel Parite Farkı'nın bir diğer dezavantajı, oranındaki farklılığın, diğer nedenlerden olabilecek etkileri ayıramamasıdır. Bu nedenle, İstatistiksel Parite Farkı tek başına eşitsizliği göstermek için yetmez ve farklı analizlere desteklenmesi gerekir [71].

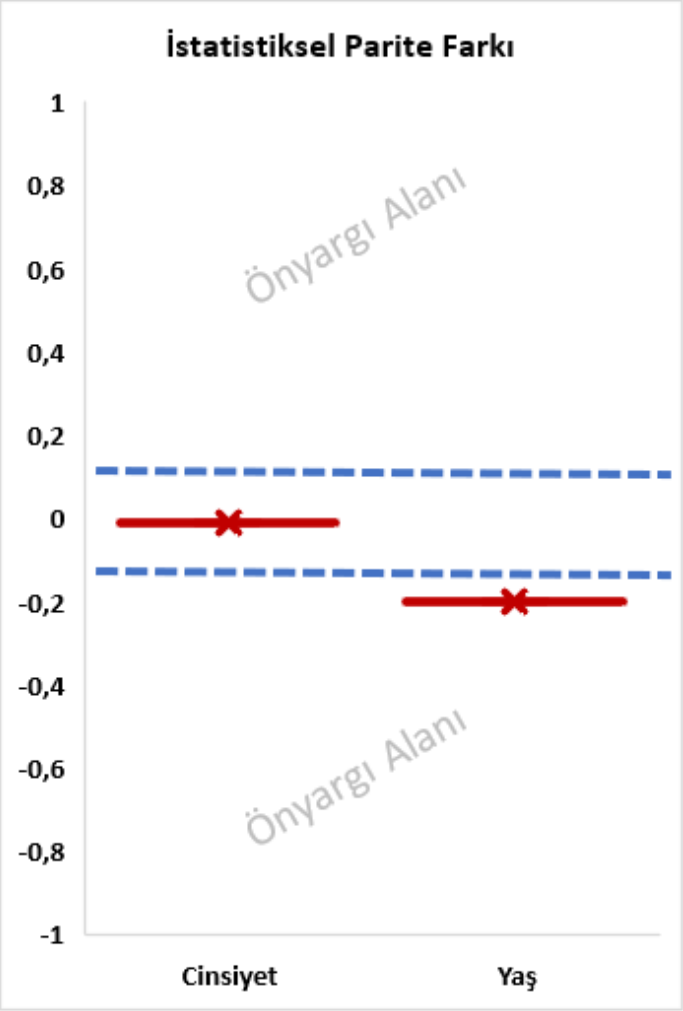
Ayrıcalıklı olmayanların ayrıcalıklı olanlara oranın pozitif ve tüm küme içindeki farkı olarak hesaplanır. Metriğin sonucu -1 ile 1 arasındadır. Algoritmik eşitsizlik olmaması için sonucunun sıfır olması gerekmektedir. Sonucun sıfırdan uzaklaşması eşitsizlik olduğu anlamına gelmektedir. Örneğin, bir işe alım sürecinde erkek adayların kabul oranı %70'iken kadın adayların kabul oranı %50 ise, İstatistiksel Parite Farkı 0.20 (yani %20) olacaktır. Bu, cinsiyet nedeniyle ayrımcılık yapıldığını gösterir. Ayrıca İstatistiksel Parite Farkının -0,1 ile 0,1 arasında çıkması kabul edilebilir seviye olarak görülmektedir [72].

$$IPF = \frac{Y(p)}{X(p)} - \frac{Y(t)}{X(t)}$$

Aşağıdaki çizelgede görüldüğü üzere yaş verisinin eşik değerlerinin üstünde olduğu ve kredi karar süreçlerinde makine öğrenmesi algoritmasının önyargı oluşturmasına başka bir ifade ile yanlamasına neden olmaktadır. Ayrıca algoritmanın cinsiyet verisi için eşitsizlik yaratmadığı görülmektedir.

Çizelge 4.2.3: İstatistiksel Parite Farkı

	Endeks Sonucu	
	Cinsiyet	Yaş
İstatistiksel Parite Farkı	-0,02	-0,3



Fark Etkisi (Disparate Impact)

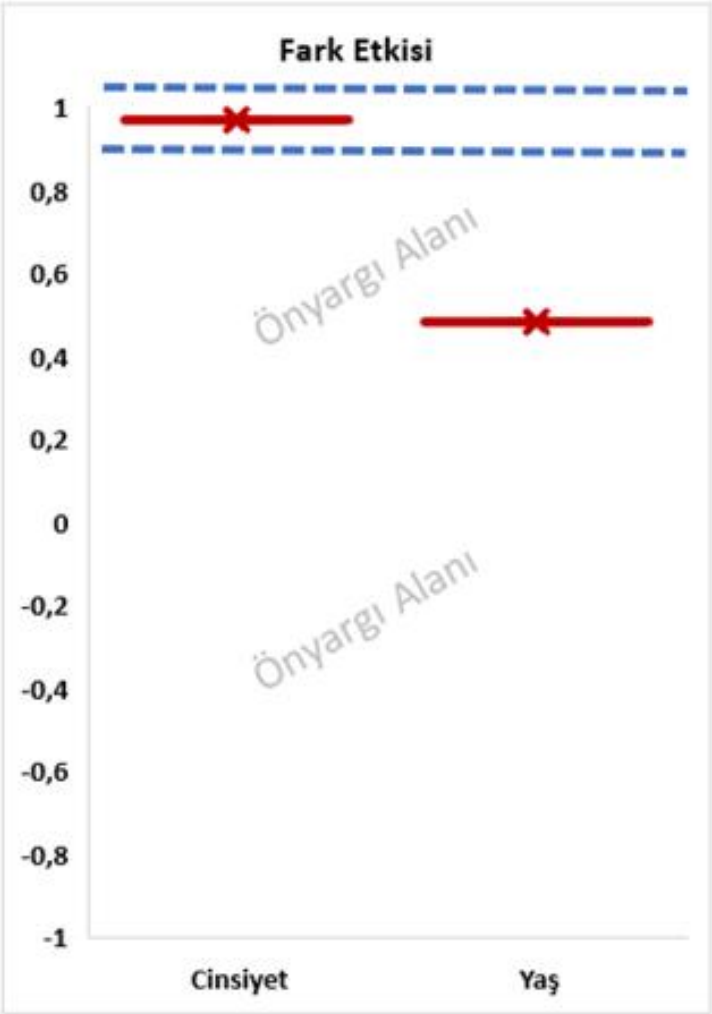
Fark Etkisi, liyakatin olmadığı işe alım süreci, terfi süreci, kredi tahsisi gibi birçok alanda sınıfsal probleme neden olabilir. Bu standart prosedürlerin dışına çıkılan uygulamalara karşı tedbirli olmalı ve eşit fırsatlar sağlamak için kontroller sağlanmalıdır. Fark Etkisi uygun olmayan eşitsizliklerin ortaya çıkmasında ölçüt olabilir. Her ikinin grubun olumlu sonuç oranının bölümü olarak hesaplanır. Metriğin sonucu sıfırdan büyüktür. Algoritmik eşitsizlik olmaması için sonucun bir olması beklenmektedir. Sonucun birden küçük olması ayrıcalıklı grup için daha yüksek fayda anlamına gelirken, birden büyük olması ise ayrıcalıksız grup için daha yüksek fayda anlamına gelmektedir. Ayrıca Fark Etkisinin 0,8 ile 1 arasında çıkması kabul edilebilir seviye olarak görülmektedir [73] [74].

$$FE = \frac{Y(p)}{Y(t)} \div \frac{X(p)}{X(t)}$$

Aşağıdaki çizelgede görüldüğü üzere yaş verisinin eşik değerlerinin altında kaldığı ve kredi karar süreçlerinde makine öğrenmesi algoritmasının önyargı oluşturmasına başka bir ifade ile yanlamasına neden olmaktadır. Ayrıca algoritmanın cinsiyet verisi için eşitsizlik yaratmadığı görülmektedir.

Çizelge 4.2.4: Fark Etkisi

	Endeks Sonucu	
	Cinsiyet	Yaş
Fark Etkisi	0,97	0,48



Fırsat Eşitliği Farkı (Equal Opportunity Difference)

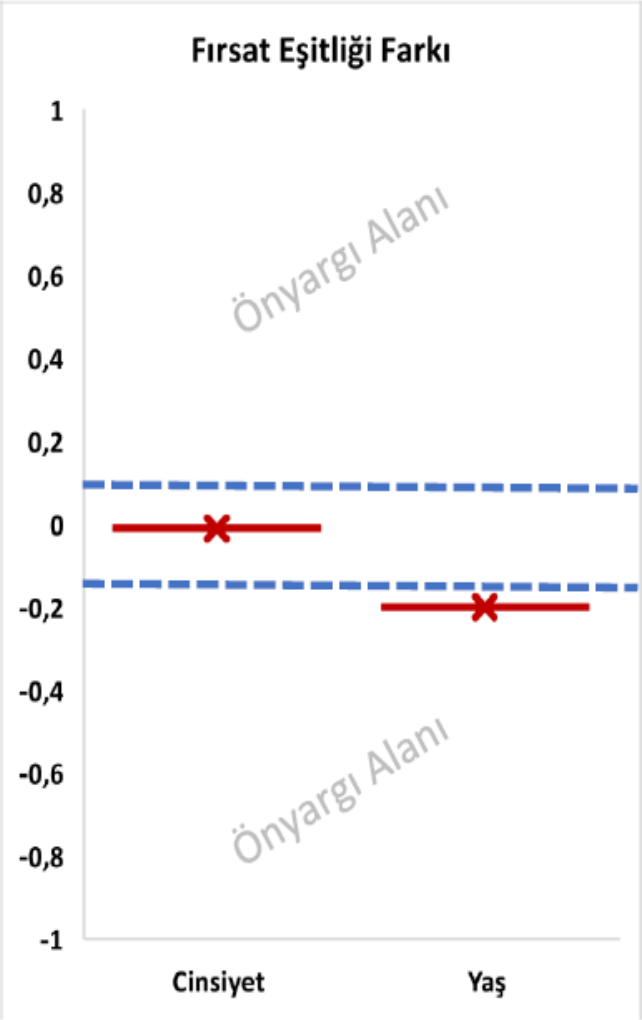
Fırsat Eşitliği Farkı, sınıfsal problemlerde ayrımcılığın tanımlanması süreçlerinde kullanılabilir. Örneğin, bir iş ilanı sadece bir cinsiyet için uygun ise, diğer cinsiyet için başvuru oranının, ilanın yayınlandığı tüm cinsiyetlerin toplam başvuru oranıyla kıyaslandığı Fırsat Eşitliği Farkı ölçütü kullanılabilir. Başka bir ifadeyle Ayrıcalıklı olmayanların ve ayrıcalıklı pozitif oranların farkı olarak hesaplanır. Metriğin sonucu -1 ile 1 arasındadır. Algoritmik eşitsizlik için sonucunun sıfır olması beklenmektedir. Sonucun sıfırdan uzaklaşması eşitsizlik olduğu anlamına gelmektedir. Aynı zamanda sonucun sıfırdan küçük olması ayrıcalıklı grup için daha yüksek fayda anlamına gelirken sıfırdan büyük olması ayrıcalıksız grup için daha yüksek fayda anlamına gelmektedir. Ayrıca fırsat eşitliği farkının -0,1 ile 0,1 arasında çıkması kabul edilebilir seviye olarak görülmektedir [75] [76].

$$FE = \frac{Y(p)}{Y(t)} \div \frac{X(p)}{X(t)}$$

Aşağıdaki çizelgede görüldüğü üzere yaş verisinin eşik değerlerinin üstünde olduğu ve kredi karar süreçlerinde makine öğrenmesi algoritmasının önyargı oluşturmasına başka bir ifade ile yanlamasına neden olmaktadır. Ayrıca algoritmanın cinsiyet verisi için eşitsizlik yaratmadığı görülmektedir.

Çizelge 4.2.5: Fırsat Eşitliği Farkı

	Endeks Sonucu	
	Cinsiyet	Yaş
Fırsat Eşitliği Farkı	-0,07	-0,43



Ortalama Oran Farkı (Average Odds Difference)

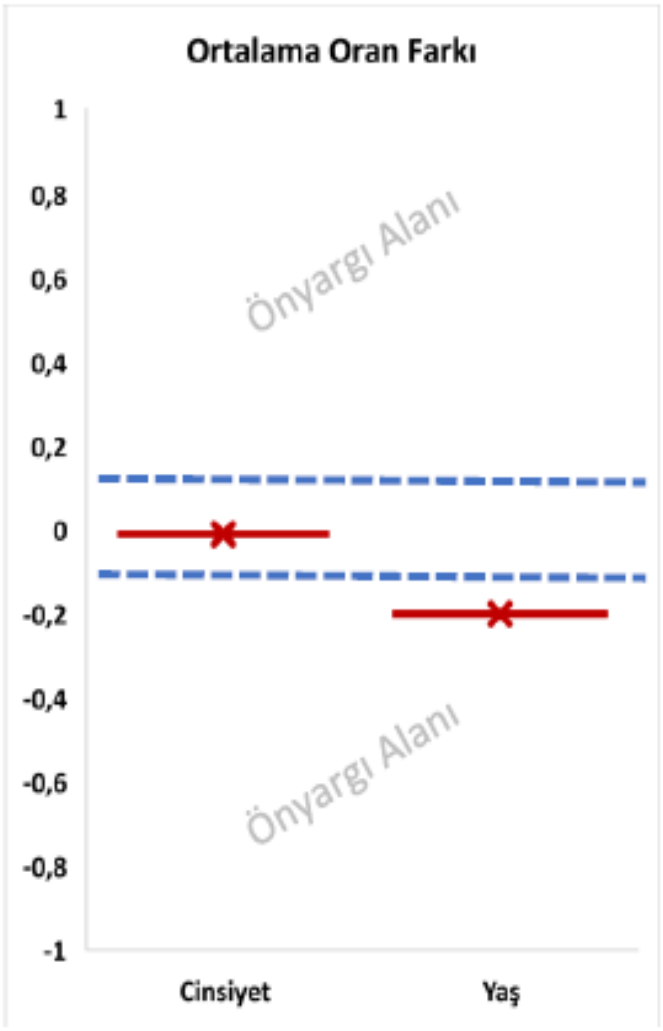
Bu metrik, ayrıcalıklı olmayan ve ayrıcalıklı gruplar arasındaki gerçek yanlış pozitif oranları ile gerçek pozitif oranının farkı olarak hesaplanır. Örneğin, bir işe alım kabul sürecinde, belirli yaş grubu için Ortalama Oran Farkı yüksekse, diğer grup için ilgili pozisyon hakkında gerekli niteliklere sahip olmadığı anlamına gelmez. Bu durum ayrımcılık yapıldığına dair referans gösterilebilir. Metriğin sonucu -1 ile 1 arasındadır. Algoritmik eşitsizlik için sonucunun sıfır olması beklenmektedir. Sonucun sıfırdan uzaklaşması eşitsizlik olduğu anlamına gelmektedir. Aynı zamanda sonucun sıfırdan küçük olması ayrıcalıklı grup için daha yüksek fayda anlamına gelirken sıfırdan büyük olması ayrıcalıksız grup için daha yüksek fayda anlamına gelmektedir. Ayrıca fırsat eşitliği farkının -0,1 ile 0,1 arasında çıkması kabul edilebilir seviye olarak görülmektedir [76] [77].

$$OOF = \frac{Y(p)}{Y(t)} - \frac{X(p)}{X(t)}$$

Aşağıdaki çizelgede görüldüğü üzere yaş verisinin eşik değerlerinin üstünde olduğu ve kredi karar süreçlerinde makine öğrenmesi algoritmasının önyargı oluşturmasına başka bir ifade ile yanlamasına neden olmaktadır. Ayrıca algoritmanın cinsiyet verisi için eşitsizlik yaratmadığı görülmektedir.

Çizelge 4.2.6: Ortalama Oran Farkı

	Endeks Sonucu	
	Cinsiyet	Yaş
Ortalama Oran Farkı	-0,01	-0,2



Eşitsizlik metriklerinin her iki veri için toplu ve detay açıklamaları aşağıda çizelge 4.1.1 ve çizelge 4.1.2'de yer almaktadır.

		Endeks Sonucu
Eşitsizlik Metrikleri	**Detay**	**Cinsiyet**
İstatistiksel Parite Farkı	Kredi Verilen Kadınların, kredi verilen erkeklere oranı eksi tüm kadınların tüm erkeklere oranı	-0,02
Fark Etkisi	Kredi verilen kadınların toplam kadınlar içindeki oranı bölü Kredi verilen erkeklerin toplam erkekler içindeki oranı	0,97
Fırsat Eşitliği Farkı	Kredi verilen kadınlar eksi kredi verilen erkekler bölü tüm insanlar	-0,07
Ortalama Oran Farkı	Kredi verilen kadınların toplam kadınlar içindeki oranı eksi Kredi verilen erkeklerin toplam erkekler içindeki oranı	-0,01

Şekil 4.1.1: Eşitsizlik Metrikleri Toplu (Cinsiyet)

		Endeks Sonucu
Eşitsizlik Metrikleri	**Detay**	**Yaş**
İstatistiksel Parite Farkı	Ortalama yaşın altında kredi verilenlerin, ortalama üstünde kredi verilenlere oranı eksi ortalama yaşın altında olanların yaş ortalamasının üstünde olanlara oranı	-0,3
Fark Etkisi	Kredi verilen ortalama yaşın altındaki kişilerin oranı bölü Kredi verilen ortalama yaşın üstündeki kişilerin oranı	0,48
Fırsat Eşitliği Farkı	Ortalama yaşın altında kredi verilenler eksi ortalama yaşın üstünde kredi verilenler bölü tüm insanlar	-0,43
Ortalama Oran Farkı	Kredi verilen ortalama yaşın altındaki kişilerin oranı eksi Kredi verilen ortalama yaşın üstündeki kişilerin oranı	-0,2

Şekil 4.1.2: Eşitsizlik Metrikleri Toplu (Yaş)

5. BULGULAR

Makine öğrenmesi destekli kredi karar süreçlerinde algoritmalar kredi başvuru sahiplerinin yaş, cinsiyet, ırk gibi demografik özellikleri de dahil olmak üzere birçok farklı özelliklerini dikkate alarak karar verirler. Ancak, bazı özelliklerin diğerlerinden daha fazla önem kazanması, ayrımcılık ve eşitsizliklerin oluşmasına neden olabilir. Bu kapsamda veri analizi bölümünde cinsiyet ve yaş değişkeni için sırasıyla İstatistiksel Parite Farkı, Fark Etkisi, Fırsat Eşitliği Farkı ve Ortalama Oran Farkı metriklerinin sonuçlarına yer verilmiştir. Yapılan nicel analizler sonrası kredi karar sürecinde kullanılan Yapay Zekâ algoritmalarının özellikle yaş özellikleri bakımında eşitsizliğe sebep olduğu tespit edilmiştir.

Aşağıda ilgili metrikler için tespit edilen bulgular sırasıyla paylaşılmaktadır.

- İstatistiksel Parite Farkı sonucuna göre, yaş özelliği bakımından onaylanma oranları arasında istatistiksel olarak anlamlı bir fark olduğu tespit edilmiştir. Bu bulgu, farklı yaş grupları arasında kabul edilme oranlarının dengesiz olduğunu göstermektedir.

- Fark Etkisi sonucuna göre, yaş gruplarındaki başvuru sahiplerinin onaylanma oranlarında mevcut başvuru durumuna göre farklılık olduğu tespit edilmiştir. Bu bulgu, farklı yaş grupları arasında başvuru sonucuna göre kabul edilme oranlarının dengesiz olduğunu göstermektedir

- Fırsat Eşitliği Farkı sonucuna göre, farklı yaş gruplarındaki başvuru sahiplerinin kabul edilme oranları arasında farklılıklar olduğu tespit edilmiştir. Bu bulgu, yaş özelliği bakımından kabul edilme oranlarının dengesiz olduğunu ve bazı yaş gruplarının diğerlerine göre daha fazla kabul edildiğini göstermektedir.

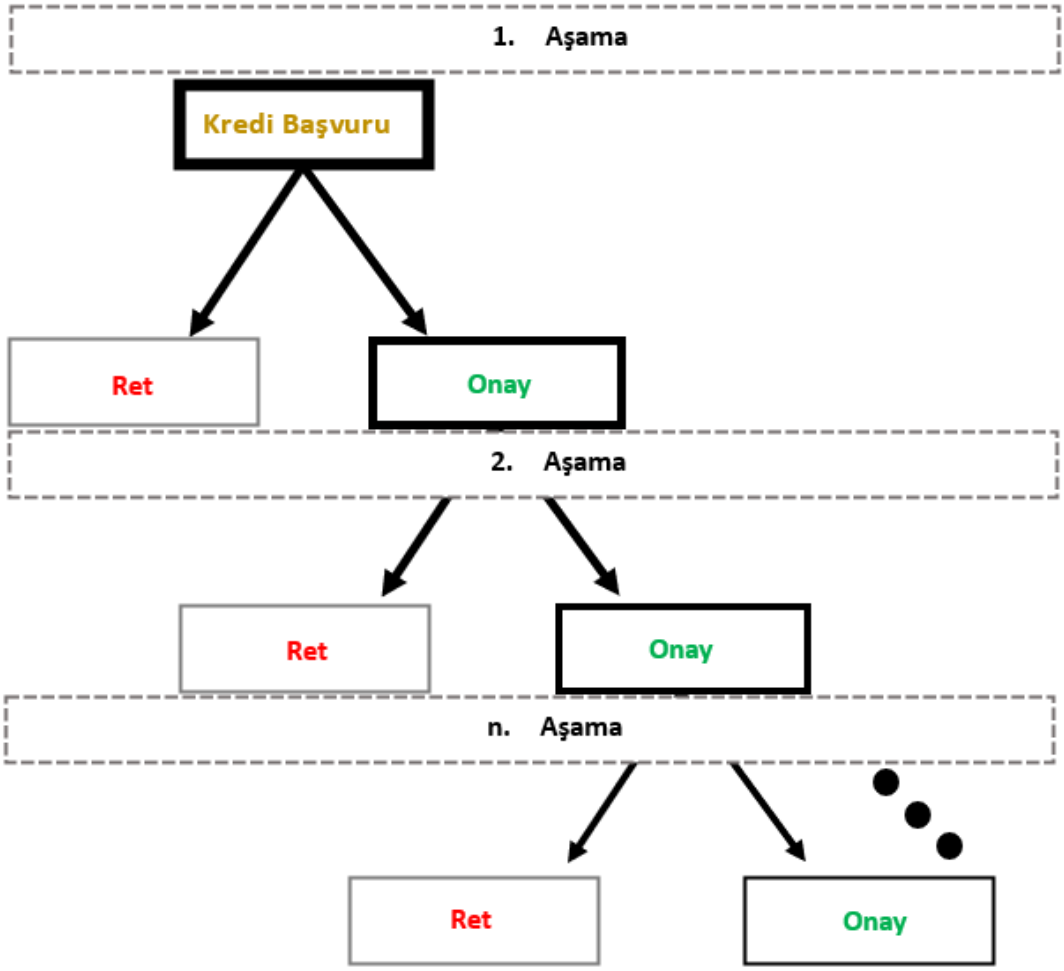
- Ortalama Oran Farkı sonucuna göre, farklı yaş gruplarındaki başvuru sahiplerinin kabul edilme ve reddedilme oranları arasında farklılıklar olduğu tespit edilmiştir. Bu bulgu, yaş özelliği nedeniyle kabul edilme ve reddedilme oranları arasındaki farkın anlamlı olduğunu gösterir ve bu da yaş özelliği bakımından ayrımcılığın varlığını göstermektedir.

Bu bulgular, kredi kararlarının demografik özelliklere dayalı olarak sonuçlarının yanlılık gösterdiği ve bu nedenle bazı grupların diğerlerine göre ayrımcılığa maruz kaldığı anlamına gelmektedir.

Yukarıda tespit edilen bulgular, makine öğrenmesi destekli modellerin kredi karar sürecinde kullanımı sırasında yaş özelliği nedeniyle eşitsizlik ve ayrımcılık yaşandığını göstermiştir. Bu durum, modellerin belirli özellikleri diğerlerinden daha fazla dikkate alması ve bu nedenle yanlılık oluşması ile açıklanabilir.

Ancak, bu modellerin sürekli olarak kullanılması ve belirli özellikleri diğerlerinden daha fazla dikkate alması sonucunda, yanlılık oranının artması ve eşitsizliğin daha da büyümesi beklenir. Özellikle, bu modellerin sonsuz defa kullanılması durumunda, çıkan sonuçların tek tip olacağı ve sadece bir tarafı seçeceği açıktır.

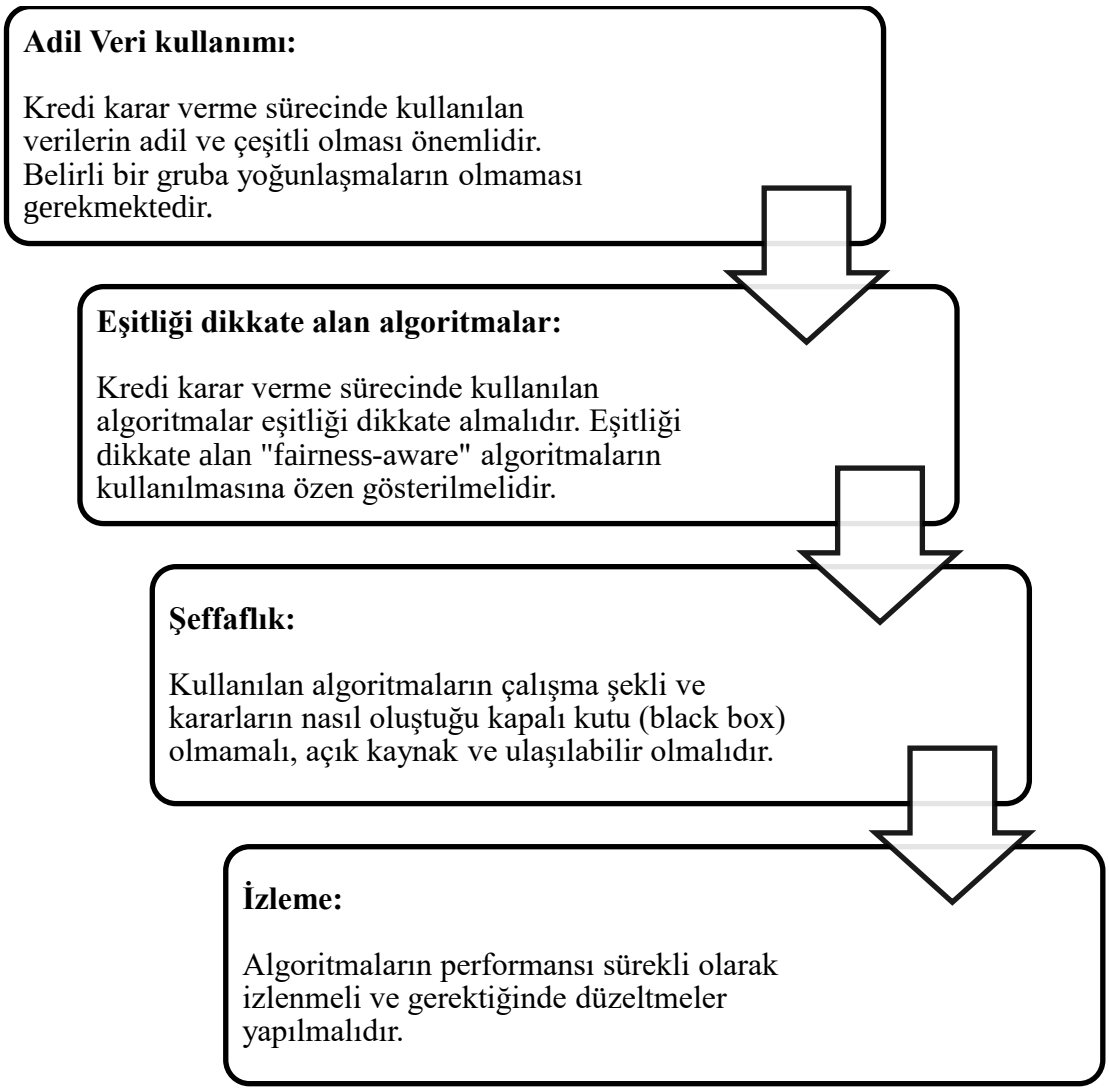
Bu durumu aşağıdaki şekil 5.1 yardımı ile görselleştirecek olursak her aşama sonunda onay kümesi ile artan doğru gözlem sayısının yaş ortalamasının yukarı çekmesine neden olacaktır. Bu durum sonucunda makine öğrenmesi destekli kredi karar desteği uygulamalarının belirli bir süre sonra gençlere kredi veremeyeceği çıkarımını yapmak mümkündür.



Şekil 5.1: Karar Ağacı Dallanması

Bu durum sonucunda, eşitsizlikler daha da artacak ve toplumda adalet duygusunun sarsılmasına neden olabileceği değerlendirilmektedir. Bu nedenle, makine öğrenmesi destekli modellerin kullanımı sırasında, özellikle demografik özelliklerin de dikkate alındığı kredi karar süreçlerinde, ayrımcılık ve yanlılık durumlarının önlenmesi için birtakım önlemler alınması gerektiği düşüncesindeyiz.

Bu önlemler, modellerin tasarımı ve veri seti hazırlığından başlayarak, sürecin her aşamasında dikkate alınmalıdır. Ayrıca, bu tür algoritmaların sonuçlarının sık sık kontrol edilmesi ve değerlendirilmesi gerekmektedir. Bu durum şekil 5.2 üzerinde özetlenmiştir.



Şekil 5.2: Yapay Zekâ Açıklama ve Kontrol Hiyerarşisi (XAI)

6. SONUÇ VE DEĞERLENDİRME

Önceki bölümlerde konusu geçen eşitsizlik metriklerin her biri, kredi karar sürecinde yaş ve benzeri verilerin özelliğinden dolayı olası ayrımcılık durumlarını tespit etmek için kullanılabilir. Ancak, ilgili metriklerin yüksek değerler göstermesi, yaş özelliğinden dolayı ayrımcılık olduğunu gösterirken sadece yaş özelliğinden dolayı ayrımcılık var olduğuna kesin kanıt olarak gösterilemez. Başta korelasyon olmak üzere diğer faktörler de değerlendirilmeli ve dikkate alınmalıdır. Makine öğrenmesi destekli kredi karar süreçlerinde kullanılan algoritmalar, bazı durumlarda eşitsizliklerin oluşmasına neden olabileceği dikkate alınarak eşit ve adil olma kapsamında ilkesel bir yaklaşım benimsenmesi önerilmektedir oluşturulmalıdır.

Makine öğrenmesi tabanlı kredi karar destek mekanizmaları, finansal kuruluşlar tarafından (banka, faktöring vb.) gerçek ve tüzel kişilere kredi verebilmek üzere tasarlanan bir çeşit karar destek sistemidir. Kredi karar sistemlerinin temelinde başvuru sahibinin krediye uygunluğunun değerlendirilmesi hedeflenir. Bu hedef doğrultusunda gerekli kontrollerin düzenli yapılması ve modelin sonuçlarının geri yönelik test (backtesting) ve doğrulama (validation) süreçleri ile değerlendirilmesi hem toplumsal hem de işletmenin faydasına olacaktır. Ayrıca bu kontroller finansal kuruluşların müşterileri arasında adalet ve eşitlik sağlamak için önemli bir adımdır.

Makine öğrenimi tabanlı kredi karar destek süreçleri günümüzde kredi değerlendirmesini otomatikleştirmek ve hızlandırmak için tercih edilen bir yöntemdir. Ancak, makine öğrenimi sistemleri çalışma prensibi geçmiş verilere dayanır ve bu veriler cinsiyet, yaş, etnik köken, eğitim seviyesi gibi faktörleri yansıtabilir. Bu nedenle, makine öğrenimi sistemleri eşitsizliği perçinleyebildiği bulgular bölümünde ve literatürde yer alan çalışmalarda görüşmüştür [78] [79]. Bu kapsamda cinsiyet eşitsizliği ve yaş problemi, kredi kararlarında sıklıkla öne çıkan durumlardır [80] [81] [82]. Bu durum veri kümesinde kadınlara ait geçmiş kredi verilerinin erkeklere göre daha az olmasından kaynaklı olabilirken toplumsal formlardan dolayı kadınların daha az iş sahibi, daha az ekonomik özgürlüğe ve daha az finansal okur yazarlığa sahip olmalarından kaynaklı da ortaya çıkabilmektedir.

Bir diğer açıklayıcı değişken olan yaş faktörünün de eşitsizliğe neden olduğu ve kredi veren kuruluşların genellikle genç bireylere kredi geçmişi ve ödeme alışkanlıkları hakkında görece daha az bilgiye sahip olmasından dolayı daha yüksek risk algısı oluşturmasına neden olabilmektedir. Cinsiyet ve yaş gibi veriler kredi karar süreçlerinde adil ve eşit bir şekilde dikkate alınmalıdır ve diğer faktörlerle birlikte değerlendirilmelidir. Bu nedenle, kredi süreçlerinde makine öğrenimi kullanılırken, eşitsizlik risklerini azaltmak için özenli bir şekilde tasarlanması, izlenmesi ve kontrol edilmesi gerekmektedir. Eşitlik ve adalet ilkelerinin korunması için etik kuralların belirlenmesi ve uygulanması önemlidir.

Makine öğrenimi tabanlı karar destek sistemlerinde algoritmik önyargı ve eşitsizlik, algoritma kararlarının belirli gruplar ya da bireyler için dezavantaj veya avantaj ile sonuçlanmasına neden olmaktadır. Algoritmik eşitsizlikleri önlemek için, veri kaynaklarının daha çeşitli ve adil hale getirilmesi, öğrenme modellerinin eşitliğe daha fazla özen gösterilmesi ve sonuçların etkililiği ve eşitliği için sürekli olarak izlenme yapılması gerekmektedir. Karar destek sistemlerinde eşitsizliğin giderilmesi ve sonuçlarının performansının artırılması için farklı yöntemlere başvurulabilir.

Muhtemel algoritmik eşitsizlikleri izlemek ve bu konuda önlemler alabilmek için yukarıda önerilen süreçlerin Yapay Zekâ sistemleri yardımıyla kullanılması mümkündür. Böylece, kullanıcılar veya karar sistemi yetkilileri, “makine öğrenimi kararlarının”, bir başka şekilde ifade etmek gerekirse, “algoritmik kararların”, hangi nedenlere dayanılarak verildiğini daha iyi bir şekilde anlamlandırılabilir.

Bu süreç Açıklanabilir Yapay Zekâ (Explainable AI- XAI) kavramına referans gösterilebilir [83] [84]. XAI konusunun ele alındığı kısıtlı sayıdaki akademik çalışmalar farklı tanımlar referans verilebilirken, genel olarak XAI aşağıdaki şekillerde özetlemek mümkündür [85] [86] [87] [88].

Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence-AYP-XAI), Yapay Zekâ sistemlerinin çalışmalarını, sonuçlarını ve kararlarını kullanıcılara ve geliştiricilere daha açık ve anlaşılabilir bir şekilde açıklamaya odaklanan bir araştırma/çalışma alanıdır.

Açıklanabilir Yapay Zekâ (AYP- XAI), Yapay Zekâ, makine öğrenmesi ve arkasındaki kullanılan algoritmaların, aldığı kararları ve sonuçları insanlar tarafından anlaşılabilir hale getirmek hedefler. Bu sayede, insanlar Yapay Zekâ sistemlerini anlamlandırabilmesini sağlar.

Açıklanabilir Yapay Zekâ (AYP- XAI), makine öğrenimi tabanlı karar verme sürecini anlamak ve açıklamak için tasarlanmış bir kavramdır. Bu sistemler, kullanıcıların veya sistemin yetkililerinin, sistemin nasıl karar verdiğini ve neden böyle karar verdiğini anlamasına olanak tanır. Açıklanabilir Yapay Zekâ sistemleri, karar destek sistemlerini düzenli izleme, performansı, sonuçları açıklama ve benzeri yöntemlerle derinlemesine analiz yaparak doğruluğunu ve eşitliğinin artırılmasını hedefler.

Yeni teknolojiler üretmekle sorumlu ABD Savunma Bakanlığı tarafından desteklenen İleri Araştırma Savunma Projeleri Araştırma Ajansı (The Defense Advanced Research Projects Agency- DARPA) DARPA, XAI için aşağıdaki tanımlamaları yapmıştır. DARPA tarafından desteklenen XAI programının amacı, Yapay Zekâ sistemlerinin kararlarının ve çıktılarının insanlar tarafından anlaşılmasını ve yorumlanmasını sağlamaktır. XAI teknolojisinin, savunma sanayi, sağlık, finans ve diğer sektörlerdeki Yapay Zekâ uygulamalarında kullanılabileceğini öngörmektedirler. Projenin temel amacı insanlar tarafından anlaşılması ve yorumlanması kolay Yapay Zekâ sistemleri projelerinin geliştirilmesidir [89] [90] [91] [92]. Tüm bu tanımlamalardan çıkarılabileceği üzere Açıklanabilir Yapay Zekâ, eşitsizliği önlemek için kullanılabilecek bir yöntem önerisi olabilir ama eşitsizliğin tamamı ile ortadan kaldıracak bir çözüm olamayabilir. Bu konuda araştırmalar halen devam etmektedir [93] [94] [95] [96] [97].

Algoritmik önyargı veri toplama ve analizinde, insanlar tarafından oluşturulan ön yargıların, yanlılıkların ve hataların bilgisayar sistemlerine yansıması olarak da görülebilir. Bu durum disiplinler arası bir etkileşimle ilgilidir çünkü algoritmik önyargı, birçok farklı disiplinde kullanılan Yapay Zekâ ve makine öğrenmesi teknolojileri ile ilgilidir. Algoritmik önyargı, veri bilimi, yönetişim, bilgisayar, istatistik, sosyal bilimler (psikoloji, sosyoloji) ve diğer birçok disiplinde etkileşim halindedir.

Veri toplama ve analizi konusunda bilgi sahibi olmayan bir araştırmacı, yanlış veri özelliklerini veya yanlış ölçümleri kullanarak, bir algoritmanın belirli bir grubu ayrımcı bir şekilde ele almasına veya farklı davranmasına yol açabilir. Bu durumda, araştırmacılar, istatistikçiler, veri bilimciler ve diğer uzmanlar arasındaki iş birliği, algoritmik önyargının önlenmesine yardımcı olabilir. Ayrıca, Yapay Zekâ sistemleri üzerinde çalışan mühendislerin, veri kümesindeki çeşitlilik eksikliği nedeniyle ayrımcı modeller oluşturmasıdır. Bu durum, sosyal bilimciler, psikologlar ve sosyologlar gibi disiplinler arası uzmanların iş birliği yapmasıyla önlenebilir. Algoritmik önyargının çözümlenmesi, farklı disiplinler arasındaki etkileşimlerin ve iş birliklerinin önemini vurgular. Farklı disiplinlerden uzmanların bir araya gelerek algoritmik önyargıyı tespit etmeleri, önlem almaları ve bunu önlemek için çalışmaları gerekmektedir.

KAYNAKLAR

[1] **McCarthy, J.** (1956). Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Dartmouth College.

[2] **A. M. Turing** (1950). Computing Machinery and Intelligence.

[3] **Newell, A., ve Simon, H. A.** (1958). The logic theory machine. IRE Transactions on Information Theory.

[4] **Arf C.** (1959). Makine düşünebilir mi ve nasıl düşünebilir. Halk Konferansları.

[5] **PricewaterhouseCoopers** (2023) Yeni Coo: Operasyonları Yapay Zekâ Ile Güçlendirmek. PwC. Retrieved January 2, 2023, from https://www.pwc.com.tr/yeni-coo-operasyonlari-yapay-Zekâ-ile-guclendirmek.

[6] **MemSQL** (2017, September 26). MemSQL takes machine learning models real-time. GlobeNewswire News Room. Retrieved January 2, 2023, from https://www.globenewswire.com/news-release/2017/09/26/1132800/0/en/MemSQL-Takes-Machine-Learning-Models-Real-Time.htm.

[7] **Lindsay, K.** (2018). Take the 2018 State of Artificial Intelligence (AI) in personalization survey. Adobe Blog. Retrieved January 2, 2023, from https://blog.adobe.com/en/publish/2018/02/06/take-2018-state-artificial-intelligence-ai-personalization-surve.

[8] **Gartner.** https://www.gartner.com/smarterwithgartner/top-10-trends-in-the-digital-world-of-work/.

[9] **IBM.** https://www.ibm.com/blogs/watson/2019/06/the-ibm-institute-for-business-value-releases-its-annual-c-suite-study/.

[10] **Dell.** https://www.delltechnologies.com/en-us/perspectives/the-next-data-decade/.

[11] **Infoys.** https://www.infosys.com/newsroom/press-releases/Documents/genome-research-report.pdf.

[12] **Clark, J. B.** (1891). Marshall's Principles of Economics. Political Science Quarterly, 6(1), 126-151.

[13] **Mankiw, N. G.** (2014). Principles of Economics. Cengage Learning.

[14] **Dantzig, G. B.** (1947). Maximization of a linear function of variables subject to linear inequalities. In Activity analysis of production and allocation (pp. 339-347). John Wiley ve Sons, Inc.

[15] **Dantzig, G. B.** (1963). Linear programming and extensions. Princeton University Press.

[16] **Bazaraa, M. S., Jarvis, J. J., ve Sherali, H. D.** (2013). Linear programming and network flows. John Wiley ve Sons.

[17] **Domingos, P.** (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books.

[18] **Miller, T.** (2018). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38.

[19] **Elsevier.** (n.d.). Scopus. Elsevier. Retrieved January 3, 2023, from https://www.elsevier.com/en-in/solutions/scopus.

[20] **Mendeley.** (n.d.). Retrieved January 3, 2023, from https://www.mendeley.com/search/.

[21] **Ulusal Tez Merkezi: Anasayfa. Ulusal Tez Merkezi**. (n.d.). Retrieved January 3, 2023, from https://tez.yok.gov.tr/UlusalTezMerkezi/.

[22] **Türk Dil Kurumu: Sözlük.** Türk Dil Kurumu Sözlükleri. (n.d.). Retrieved January 3, 2023, from https://sozluk.gov.tr/.

[23] **Türkiye Bilim Terimleri.** (n.d.). Retrieved January 3, 2023, from https://www.bilimterimleri.com/.

[24] **Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi.** (n.d.). Retrieved January 3, 2023, from https://cbddo.gov.tr/.

[25] **McCarthy, J.** (2004). What is artificial intelligence. URL: http://www-formal. stanford. edu/jmc/whatisai. html.

[26] **Pirim, A. G. H.** (2006). Yapay Zekâ. Yaşar Üniversitesi E-Dergisi, 1(1), 81-93.

[27] **Köroglu, Y.** (2017). Yapay Zekâ'nın teorik ve pratik sınırları. Bogaziçi Üniversitesi Yayınevi.

[28] **Adalı, E.** (2017). "Yapay Zekâ". İstanbul Teknik Üniversitesi Vakfı Yayını, Ocak-Mart, Sayı 75 https://www.ituvakif.org.tr/dergi/sayi_75.pdf (25 Mart 2019).

[29] **Sönmez, C.** (n.d.). Yapay Zekâ İcerikleri. Retrieved January 2, 2023, from https://web.itu.edu.tr/~sonmez/lisans/ai/yapay_Zekâ_icerik1_1.6.pdf.

[30] **IBM Cloud Education.** (n.d.). Yapay Zekâ (AI) nedir? IBM. Retrieved January 3, 2023, from https://www.ibm.com/tr-tr/cloud/learn/what-is-artificial-intelligence.

[31] **Russell, S. J.** (2021). Artificial Intelligence: A modern approach. Prentice Hall/Pearson Education.

[32] **Oracle Türkiye.** Yapay Zekâ (AI) nedir? Yapay Zekâ (AI) nedir? | (n.d.). Retrieved January 3, 2023, from https://www.oracle.com/tr/artificial-intelligence.

[33] **SAS.** (n.d.). Retrieved January 3, 2023, from https://www.sas.com/tr_tr/insights/analytics/yapay-Zekâ-nedir.

[34] **IBM Cloud Education.** (n.d.). Makine Öğrenmesi (ML) nedir? IBM. Retrieved January 3, 2023, from https://www.ibm.com/tr-tr/cloud/learn/machine-learning.

[35] **Microsoft Azure.** (n.d.). Retrieved January 3, 2023, from https://azure.microsoft.com/tr-tr/resources/cloud-computing-dictionary/what-is-machine-learning.

[36] **Amazon.** Retrieved January 3, 2023, from https://aws.amazon.com/tr/machine-learning/.

[37] **SAS.** Makine öğrenimi: Nedir ve neden önemlidir? Retrieved January 3, 2023, from https://www.sas.com/tr_tr/insights/analytics/machine-learning.

[38] **Oracle Türkiye.** (n.d.). Retrieved January 3, 2023, from https://www.oracle.com/tr/artificial-intelligence/machine-learning/what-is-machine-learning.

[39] **Angwin, J., Larson, J., Mattu, S., ve Kirchner, L.** (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.

[40] **Dastin, J.** (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters.

[41] **Buolamwini, J., ve Gebru, T.** (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77-91.

[42] **Sengupta, S.** (2016, December 13). When algorithms discriminate: Amazon's same-day delivery and pricing. Harvard Business Review.

[43] **Liu, Y., ve Deldjoo, Y.** (2019). Gender bias in job advertising on Facebook: An audit study. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-23.

[44] **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.

[45] **Maani, S., Han, Y., ve Laseter, T. M.** (2019). Ride-hailing surge pricing: Evidence from Uber and Lyft in metropolitan areas. International Journal of Hospitality Management.

[46] **Pessach, D., ve Shmueli, E.** (2022). A Review on Fairness in Machine Learning. ACM Computing Surveys (CSUR), 55(3), 1-44.

[47] **Barocas, S., Hardt, M., ve Narayanan, A.** (2017). Fairness in machine learning. Nips tutorial, 1, 2.

[48] **Caton, S., ve Haas, C.** (2020). Fairness in machine learning: A survey. arXiv preprint arXiv:2010.04053.

[49] **Wang, Q., Ma, Y., Zhao, K., ve Tian, Y.** (2022). A comprehensive survey of loss functions in machine learning. Annals of Data Science, 9(2), 187-212.

[50] **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.

[51] **Martínez-Plumed, F., Prudêncio, R. B., Martínez-Usó, A., ve Hernández-Orallo, J.** (2019). Item response theory in AI: Analysing machine learning classifiers at the instance level. Artificial intelligence, 271, 18-42.

[52] **Fabris, A., Messina, S., Silvello, G., ve Susto, G. A.** (2022). Tackling documentation debt: a survey on algorithmic fairness datasets. In Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-13).

[53] **Ras, G., Xie, N., van Gerven, M., ve Doran, D.** (2022). Explainable deep learning: A field guide for the uninitiated. Journal of Artificial Intelligence Research, 73, 329-397.

[54] **Pessach, D., ve Shmueli, E.** (2020). Algorithmic fairness. arXiv preprint arXiv:2001.09784.

[55] **Aziz, H., Li, B., Moulin, H., ve Wu, X.** (2022). Algorithmic fair allocation of indivisible items: A survey and new questions. arXiv preprint arXiv:2202.08713.

[56] **Wang, G., Hanashiro, R., Guha, E., ve Abernethy, J.** (2022). On Accelerated Perceptrons and Beyond. arXiv preprint arXiv:2210.09371.

[57] **Caton, S., ve Haas, C.** (2020). Fairness in machine learning: A survey. arXiv preprint arXiv:2010.04053.

[58] **Biswas, S., ve Rajan, H.** (2020, November). Do the machine learning models on a crowd sourced platform exhibit bias? an empirical study on model fairness. In Proceedings of the 28th ACM joint meeting on European software engineering conference and symposium.

[59] **Le Quy, T., Roy, A., Iosifidis, V., Zhang, W., ve Ntoutsi, E.** (2022). A survey on datasets for fairness-aware machine learning. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1452.

[60] **Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., ve Walther, A.** (2022). Predictably unequal? The effects of machine learning on credit markets. The Journal of Finance, 77(1), 5-47.

[61] **Kasmi, M. L.** (2021). Machine Learning Fairness in Finance: An Application to Credit Scoring (Doctoral dissertation, Tilburg University).

[62] **Szepannek, G., ve Lübke, K.** (2021). Facing the challenges of developing fair risk scoring models. Frontiers in artificial intelligence, 4.

[63] **Alves, G., Bernier, F., Couceiro, M., Makhlouf, K., Palamidessi, C., ve Zhioua, S. (2022).** Survey on Fairness Notions and Related Tensions. arXiv preprint arXiv:2209.13012.

[64] **Pagano, T. P., Loureiro, R. B., Araujo, M. M., Lisboa, F. V. N., Peixoto, R. M., Guimaraes, G. A. D. S., ve Nascimento, E. G. S.** (2022). Bias and unfairness in machine learning models: a systematic literature review. arXiv preprint arXiv:2202.08176.

[65] **Choudhary, M., Laclau, C., ve Largeron, C.** (2022). A Survey on Fairness for Machine Learning on Graphs. arXiv preprint arXiv:2205.05396.

[66] **Dablain, D., Krawczyk, B., ve Chawla, N.** (2022). Towards a holistic view of bias in machine learning: Bridging algorithmic fairness and imbalanced learning. arXiv preprint arXiv:2207.06084.

[67] **Fabris, A., Messina, S., Silvello, G., ve Susto, G. A.** (2022). Algorithmic Fairness Datasets: the Story so Far. arXiv preprint arXiv:2202.01711.

[68] **Raimondi, F. E., Lawrence, A. R., ve Chockler, H.** (2022). Equality of Effort via Algorithmic Recourse. arXiv preprint arXiv:2211.11892.

[69] **Liao, Y., ve Naghizadeh, P.** (2022). Social Bias Meets Data Bias: The Impacts of Labeling and Measurement Errors on Fairness Criteria. arXiv preprint arXiv:2206.00137.

[70] **Zliobaite, I.** (2017). Fairness-aware machine learning: a perspective. arXiv preprint arXiv:1708.00754.

[71] **Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., ve Venkatasubramanian, S.** (2015). Certifying and removing disparate impact. In proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 259-2.

[72] **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).

[73] **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.

[74] **Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., ve Venkatasubramanian, S.** (2015, August). Certifying and removing disparate impact. In proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 259-2.

[75] **Hardt, M., Price, E., ve Srebro, N.** (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 3325-3333.

[76] **Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ve Zhang, Y.** (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM Journal of Research and Development, 63(4/5), 4-1.

[77] **Chouldechova, A.** (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153-163.

[78] **Hurlin, C., Pérignon, C., ve Saurin, S.** (2022). The fairness of credit scoring models. arXiv preprint arXiv:2205.10200.

[79] **Jammalamadaka, K. R., ve Itapu, S.** (2022). Responsible AI in automated credit scoring systems. AI and Ethics, 1-11.

[80] **Dietrich, J.** (2005). Searching for age and gender discrimination in mortgage lending. Comptroller of the Currency.

[81] **Bacha, S., ve Azouzi, M. A.** (2019). How gender and emotions bias the credit decision-making in banking firms. Journal of Behavioral and Experimental Finance, 22, 183-191.

[82] **Avery, R. B., Brevoort, K. P., ve Canner, G.** (2012). Does Credit Scoring Produce a Disparate Impact? Real Estate Economics, 40, S65-S114.

[83] **Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ve Herrera, F.** (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 8.

[84] **Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., ve Castignani, G.** (2022). Explainable Artificial Intelligence (XAI) in Insurance. Risks, 10(12), 230.

[85] **Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., ve Yang, G. Z.** (2019). XAI—Explainable artificial intelligence. Science robotics, 4(37), eaay7120.

[86] **Gunning, D., ve Aha, D.** (2019). DARPA's explainable artificial intelligence (XAI) program. AI magazine, 40(2), 44-58.

[87] **Čyras, K., Rago, A., Albini, E., Baroni, P., ve Toni, F.** (2021). Argumentative XAI: a survey. arXiv preprint arXiv:2105.11266.

[88] **Gunning, D.** (2017). Explainable artificial intelligence (xai). Defense advanced research projects agency (DARPA), nd Web, 2(2), 1.

[89] **Gunning, D., ve Aha, D.** (2019). DARPA's explainable artificial intelligence (XAI) program. AI magazine, 40(2), 44-58.

[90] **Gunning, D., Vorm, E., Wang, Y., ve Turek, M.** (2021). DARPA's explainable AI (XAI) program: A retrospective. Authorea Preprints.

[91] **Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., ve Zhu, J.** (2019). Explainable AI: A brief survey on history, research areas, approaches and challenges. In Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, D.

[92] **Hu, B., Tunison, P., Vasu, B., Menon, N., Collins, R., ve Hoogs, A.** (2021). XAITK: The explainable AI toolkit. Applied AI Letters, 2(4), e40.

[93] **Gerlings, J., Shollo, A., ve Constantiou, I.** (2020). Reviewing the need for explainable artificial intelligence (xAI). arXiv preprint arXiv:2012.01007.

[94] **Förster, M., Klier, M., Kluge, K., ve Sigler, I.** (2020). Evaluating explainable Artifical intelligence–What users really appreciate.

[95] **Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., ve Baum, K.** (2021). What do we want from Explainable Artificial Intelligence (XAI) – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI researc.

[96] **Ignatiev, A.** (2020, July). Towards Trustable Explainable AI. In IJCAI (pp. 5154-5158).

[97] **Şengöz, N., ve Yiğit, T.** (2022, September). Towards Third Generation AI: Explainable and Interpretable AI. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 523-526). IEEE.

[98] **Taylor, P.** (2022, September 8). Total Data Volume Worldwide 2010-2025. Statista. Retrieved January 2, 2023, from https://www.statista.com/statistics/871513/worldwide-data-created/.

[99] **Chen, H., Chiang, R. H., ve Storey, V. C.** (2012). "Business Intelligence and Analytics: From Big Data to Big Impact." MIS Quarterly, 36(4), [1165-1188].

[100] **Borgman, C. L.** (2015). "Big Data, Little Data, No Data: Scholarship in the Networked World." MIT Press.

[101] **Hilbert, M., ve López, P.** (2011). "The world's technological capacity to store, communicate, and compute information." Science, 332(6025), 60-65.

[102] **Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., ve Byers, A. H.** (2011). "Big data"