
YÖNETİM BİLİŞİM SİSTEMLERİNDE İLERİ ARAŞTIRMALAR

Editör: Doç.Dr. Mahmut TOKMAK

Yönetim Bilişim Sistemlerinde İleri Araştırmalar

Editör

Doç. Dr. Mahmut TOKMAK

yaz
yayınları

2024

Yönetim Bilişim Sistemlerinde İleri Araştırmalar

Editör: Doç. Dr. Mahmut TOKMAK

© YAZ Yayınları

Bu kitabın her türlü yayın hakkı Yaz Yayınları'na aittir, tüm hakları saklıdır. Kitabın tamamı ya da bir kısmı 5846 sayılı Kanun'un hükümlerine göre, kitabı yayınlayan firmanın önceden izni alınmaksızın elektronik, mekanik, fotokopi ya da herhangi bir kayıt sistemiyle çoğaltılamaz, yayınlanamaz, depolanamaz.

E_ISBN 978-625-5547-03-3

Aralık 2024 – Afyonkarahisar

Dizgi/Mizanpaj: YAZ Yayınları

Kapak Tasarım: YAZ Yayınları

YAZ Yayınları. Yayıncı Sertifika No: 73086

M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar/AFYONKARAHİSAR

www.yazyayinlari.com

yazyayinlari@gmail.com

info@yazyayinlari.com

İÇİNDEKİLER

Örgütlerde Sezgisel ve Rasyonel Karar Vermede Yapay Zekânın İşlevi	1
---	----------

Hüseyin ŞATIRER, Osman DABAN, Ahmet Sait ÖZKUL

Veri Madenciliği Araçlarına Genel Bir Bakış.....	13
---	-----------

Mahmut TOKMAK

Büyük Veri, Veri Analitiği ve İşletmeler Açısından Önemi	29
---	-----------

Levent KOŞAN

"Bu kitapta yer alan bölümlerde kullanılan kaynakların, görüşlerin, bulguların, sonuçların, tablo, şekil, resim ve her türlü içeriğin sorumluluğu yazar veya yazarlarına ait olup ulusal ve uluslararası telif haklarına konu olabilecek mali ve hukuki sorumluluk da yazarlara aittir."

ÖRGÜTLERDE SEZGİSEL VE RASYONEL KARAR VERMEDE YAPAY ZEKÂNIN İŞLEVİ

Hüseyin ŞATIRER¹

Osman DABAN²

Ahmet Sait ÖZKUL³

1. GİRİŞ

Karar verme süreçleri, örgütlerin etkinlik ve verimliliğini belirleyen en kritik unsurlar arasında yer almakta ve çoğu zaman karmaşık yapısıyla da dikkat çekmektedir. Geleneksel olarak sezgisel ve rasyonel yaklaşımlarla çözülmeye çalışılan bu süreçler, yapay zekâ teknolojilerinin gelişimiyle birlikte önemli bir dönüşüme uğramıştır. Artık karar alma, sadece bireysel liderlik yeteneklerine dayalı bir mekanizma olmaktan çıkıp, çok boyutlu bilişsel öğelerin ve yapay zekânın sunduğu analitik kapasitenin etkisi altında şekillenmektedir. Örgütlerin karşılaştığı karmaşık problemler ve hızla değişen çevre koşulları, sezgisel ve rasyonel yaklaşımların dengeli bir şekilde kullanılmasını gerektirmektedir.

Sezgisel karar verme, yöneticilerin deneyimlerine ve içgörülerine dayanarak hızlı ve pratik çözümler geliştirmesine olanak tanırken; rasyonel karar verme ise daha sistematik ve mantıksal bir çerçevede karar alınmasını hedeflemektedir. Yapay zekâ teknolojileri, bu iki yaklaşımı destekleyerek, sezgisel karar verme süreçlerinde hız ve içgörü, rasyonel süreçlerde ise derinlemesine analiz ve veriye dayalı öngörüler sunmaktadır. Sınırlı rasyonalite teorisi ise yapay zekânın insan karar vericilerin bilişsel sınırlamalarını aşmadaki rolünü anlamak için önemli bir çerçeve sunmaktadır. Bu çalışma, örgütlerde sezgisel ve rasyonel karar verme süreçlerine yapay zekânın entegrasyonunu inceleyerek, bu entegrasyonun sağladığı avantajları ve beraberinde getirdiği zorlukları kapsamlı bir şekilde ele almayı amaçlamaktadır.

1 Öğr. Gör. Dr., (Süleyman Demirel Üniversitesi, Sağlık Hizmetleri MYO, Tıbbi Hizmetler ve Teknikler), huseyinsatirer@sdu.edu.tr, ORCID: 0000-0003-1334-4403.

2 Öğr. Gör., (Süleyman Demirel Üniversitesi, Sağlık Hizmetleri MYO, Tıbbi Hizmetler ve Teknikler), osmandaban@sdu.edu.tr, ORCID: 0000-0002-0743-7729.

3 Doç. Dr., (Süleyman Demirel Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme), saitozkul@sdu.edu.tr, ORCID: 0000-0001-8858-4685.

2. KARAR VERMEDE YAPAY ZEKÂNIN KULLANIMI

Yapay zekânın (Artificial Intelligence-AI) örgütsel karar alma süreçlerine entegrasyonu, çeşitli sektörlerde kararların nasıl alındığını yeniden şekillendiren dönüştürücü bir güç olarak ortaya çıkmıştır. YZ teknolojileri, veriye dayalı içgörüler sağlayarak, verimliliği artırarak ve kuruluşların karmaşık ortamlarda daha fazla çeviklikle gezinmesini sağlayarak karar verme sürecini geliştirmektedir. Bu bağlamda, YZ'nin örgütsel karar alma sürecindeki çok yönlü rolü ortaya konularak; faydaları, zorlukları ve gelecekteki uygulamalar için çıkarımları dikkate sunulmaktadır.

YZ'nin büyük miktarda veriyi işleme ve kalıpları belirleme kapasitesi, karar verme sürecindeki en önemli avantajlarından biridir. Örneğin Olanıyan (2023), örgütlerin eğilimleri tahmin etmesine ve işgücü yönetimini optimize etmesine olanak tanıyan, stratejik insan kaynakları karar verme sürecini geliştirmede yapay zekâ odaklı yetenek analitiğinin rolünü vurgulamaktadır. Benzer şekilde Volosova ve Матюхина (2020), YZ'nin yatırımları nasıl tahmin edebileceğini ve risk faktörlerini nasıl değerlendirebileceğini, böylece belirsizlik koşulları altında örgütsel yönetişimi nasıl destekleyebileceğini tartışmaktadırlar. Bu noktada yeteneklerin, örgütlerin daha iyi performans ve rekabet avantajı sağlayabilecek daha bilinçli kararlar alınmasını sağlayacağı belirtilebilir. Aynı zamanda YZ'nin insan karar verme süreçlerini destekleme yeteneği çeşitli bağlamlarda çok önemlidir. Örneğin Brown ve arkadaşları (2023), risklerin özellikle yüksek olduğu klinik karar verme süreçlerinde kullanılan YZ sistemlerinde şeffaflığın öneme dikkat çekmektedir. Bu şeffaflık ihtiyacı, YZ-insan işbirliğinin karar verme süreçlerinde prosedürel adalet algılarını nasıl etkilediğini araştırarak Jiang ve arkadaşlarının (2023) çalışmalarında da belirtilmektedir. Böylece YZ'nin karar verme süreçlerine dahil edilmesi, yalnızca verilen/verilecek kararların kalitesini artırmakla kalmamakta, aynı zamanda Gualdi ve Cordella (2021) tarafından tartışıldığı gibi hesap verebilirlik ve şeffaflıkla ilgili etik hususları da ele alarak dikkatlere sunmaktadır.

YZ'nin karar verme süreçlerinde uygulanmasının sağladığı kolaylıklarla birlikte zorluklarından da bahsedilmelidir. Örneğin etik çıkarımlar, YZ algoritmalarındaki önyargılar ve iş değiştirme potansiyeli, örgütlerin üstesinden gelmesi gereken önemli sorunlardandır. Balbaa (2024), bu zorlukların kapsamlı bir incelemesini sunmakta ve örgütlerin gelişmiş karar verme için YZ'den yararlanırken etik hususları ele alma ihtiyacına dikkat çekmektedir. Ayrıca, YZ sistemleri tarafından üretilen yanlış bilgi potansiyeli, Huy (2023)'un belirttiği üzere karar verme davranışı için risk oluşturmaktadır. Bu nedenle örgütler, YZ teknolojilerinin sorumluluk bilinci ekseninde kullanılmasını sağlayan çerçeveler geliştirmelidirler.

YZ'nin gelişimiyle birlikte, örgüt içindeki geleneksel karar alma yapılarının yeniden değerlendirilmesi de gereklidir. Shrestha ve arkadaşları (2019), insan ve YZ tabanlı karar verme süreçlerini yorumlanabilirlik ve karar verme hızı gibi çeşitli boyutlar açısından karşılaştıran bir çerçeve önererek, hibrit bir yaklaşımın en iyi sonuçları verebileceğini öne sürmektedirler. Bu bakış açısı, aşırı bilgi yüküyle mücadele etmek için YZ'yi yönetsel bilgi işlemeye entegre etmenin önemini vurgulayan Dietzmann ve Duan'ın (2022) bulgularıyla uyumludur. Bu anlayışla örgütler, karar verme süreçlerini YZ'yi içerecek şekilde yeniden tanımlayarak, hızla değişen bir ortamda uyarlanabilirliklerini ve yanıt verebilirliklerini artırabilirler. Bunların yanı sıra YZ, veri analizini geliştirerek, verimliliği artırarak ve etik hususları ele alarak örgütsel karar vermede oldukça önemli bir rol oynamaktadır. YZ'nin sağladığı faydalar önemli olsa da örgütler uygulamayla ilgili zorluklar konusunda da dikkatli olmalıdırlar.

3. SEZGİSEL KARAR VERMEDE YAPAY ZEKÂ

Sezgisel karar verme, psikoloji, ekonomi ve yapay zekâ (AI) dahil olmak üzere çeşitli alanlarla kesişen kritik bir çalışma alanıdır. Sezgisel yöntemler, karar verme süreçlerini basitleştiren ve bireylerin kapsamlı bir şekilde düşünmelerine gerek kalmadan hızlı bir şekilde karar vermelerini sağlayan zihinsel kısayollardır. Bu yaklaşım, özellikle geleneksel rasyonel karar verme modellerinin bilişsel aşırı yük(ü)me veya zaman kısıtlamaları nedeniyle başarısız olabileceği belirsizlik ve karmaşıklık ile karakterize edilen ortamlarda geçerlidir. Sezgisel yöntemlerin YZ sistemlerine entegrasyonu, örgütlerin karar verme yeteneklerini daha da geliştirerek onları gerçek dünya uygulamalarında daha uyarlanabilir ve verimli hale getirebilir.

Herbert Simon tarafından ortaya atılan sınırlı rasyonellik teorisi, karar vericilerin sınırlı bilgi, bilişsel sınırlamalar ve zaman baskısı kısıtlamaları altında faaliyet gösterdiğini ve bunun da karmaşık ortamlarda gezinmek için sezgisel yöntemlerin kullanılmasını gerektirdiğini öne sürmektedir (Bahreini vd., 2022). Bu teori, sezgisel yöntemlerin hem insan hem de yapay zekâ karar verme bağlamlarında nasıl işlediğini anlamada temel teşkil etmektedir. Örneğin Bahreini ve arkadaşları (2022), kullanıcıların güvenlik karar verme sürecinde kullanılabilirlik, temsil edilebilirlik ve sabitleme gibi çeşitli sezgisel yöntemler kullandığını göstererek, görev türünün sezgisel kullanımda bir moderatör olarak rolünü vurgulamaktadırlar. Bu bulgu, bilgi güvenliğini içerenler de dahil olmak üzere farklı karar verme senaryolarında sezgisel yöntemlerin uyarlanabilirliğinin altını çizmektedir. Bununla beraber Albar ve Jetter (2009), sezgisel yöntemlerin, optimal çözümleri garanti etmese de belirsiz ortamlarda tatmin edici sonuçlar verebilecek pratik kurallar olarak hizmet ettiğini belirtmektedirler. Bu ba-

kış açısı, yapay zekâ destekli karar vermede sezgisel yöntemlerin rolünün yeniden değerlendirilmesini gerektirmekte ve rasyonel modelleri sezgisel stratejilerle birleştiren hibrit bir yaklaşımın uygulamada daha etkili olabileceğini düşündürmektedir.

Yapay zekâ alanında, sezgisel yöntemlerin makine öğrenimi ile kombinasyonu, karmaşık görevlerin optimize edilmesinde umut verici sonuçlar göstermiştir. Gerlach ve arkadaşları (2022), sezgisel arama yöntemlerini makine öğrenimi ile entegre eden YZ tabanlı optimizasyon tekniklerinin radyasyon terapisinde robotik ultrason kılavuzluğunun kurulumunu nasıl geliştirebileceğini göstermektedir. Bu sinerji yalnızca verimliliği artırmakla kalmaz, aynı zamanda yapay zekâ sistemlerinin karmaşık problem alanlarında gezinmek için sezgisel yöntemlerden yararlanarak insan benzeri karar verme süreçlerini taklit etme potansiyelini de gösterir. Bu tür uygulamalar, sistemlerin dinamik ortamlarda hızlı kararlar almasını sağladıkları için YZ’de sezgisel yöntemlerin artan önemini vurgulamaktadır.

Bununla birlikte, sezgisel yöntemlere güvenmenin tuzakları da yok değildir. Ahmad ve arkadaşları (2020), sezgisel güdümlü karar verme sürecinden kaynaklanan bilişsel önyargıların, özellikle belirsizliğin yaygın olduğu girişimcilik bağlamlarında nasıl önemli hatalara yol açabileceğini tartışmaktadırlar. Sezgisel yöntemlerin daha hızlı karar verilmesini kolaylaştırabildiği gibi, özellikle bireyler önyargılarının farkında olmadıklarında, bu kararların kalitesini de tehlikeye atabileceğini vurgulamaktadırlar. Bu ikilik, bilişsel sezgisel yöntemlerin COVID-19 salgını gibi krizler sırasında stratejik karar vermeyi olumsuz etkileyebileceğini gösteren başka bir çalışmanın bulgularında görülmektedir (Ahmad, Wu, Naveed ve Ali, 2022). Bu bulgular, sezgisel yöntemlerin karar alma hızını artırabildiği gibi, doğru yönetilmediği takdirde önemli riskleri de beraberinde getirebileceğini ortaya koymaktadır.

Sezgisel yöntemlerin uygulandığı bağlam, karar verme süreçlerinin etkililiğinde önemli bir rol oynamaktadır. Guercini (2023), sezgisel yöntemlerin kapsamının, görev ortamından etkilendiğini savunmakta ve bağlamdaki değişikliklerin karar verme süreçlerinde sezgisel kuralların uygulanabilirliğini ve başarısını önemli ölçüde değiştirebileceğini belirtmektedir. Bu kavram, özellikle dijitalleşmenin karar verme ortamını dönüştürdüğü ve pazarlamacılar tarafından kullanılan sezgisel yöntemlerin yeniden değerlendirilmesini gerektirdiği pazarlama otomasyonu ile ilgilidir (Guercini, 2022). Sezgisel yöntemlerin farklı bağlamlara uyarlanabilirliği hem insan hem de yapay zekâ karar verme çerçevelerinde geçerliliklerini sürdürmeleri için çok önemlidir. Bunlarla birlikte sezgisel yöntemlerin psikolojik temelleri, karmaşıklıklarını da işaret etmektedir. Zimmermann ve arkadaşları (2021), genetik testler gibi karmaşık karar verme senaryolarında, sezgisel yöntemlerin aşırı bilgi yükünü hafifletmek ve daha yöne-

tilebilir karar süreçlerini kolaylaştırmak için hayati araçlar olarak hizmet ettiğini belirtmektedirler. Bu durum, sezgisel yöntemlerin kararları basitleştirirken aynı zamanda en alakalı bilgilere odaklanarak kararların kalitesini de artırabileceği yönündeki anlayışla uyumludur. Basitleştirme ve doğruluk arasındaki denge, karar yorgunluğunun baskı altında nasıl daha sezgisel temelli yargılara yol açabileceğini araştıran Hirshleifer ve arkadaşlarının (2019) çalışmalarında da görüldüğü gibi, sezgisel araştırmalarda tekrar eden bir temadır.

Sezgisel karar vermenin etkileri, bireysel seçimlerin ötesine geçerek örgütsel bağlamlara da uzanmaktadır. Örneğin, Riva ve arkadaşları (2011) tarafından yapılan çalışmada, hızlı ve tutumlu sezgisel yöntemlerin tıp uzmanlarının kapsamlı veri analizi yapmadan etkili kararlar almalarına nasıl yardımcı olabileceği vurgulanmaktadır. Bu yaklaşım özellikle zamanında karar vermenin kritik olduğu yüksek riskli ortamlarda faydalıdır. Benzer şekilde Banks ve arkadaşları (2020) da askeri karar alma süreçlerinde hızlı ve tutumlu sezgisel yöntemlerin faydasını vurgulamakta ve deneyimli karar alıcıların operasyonel etkinliği artırmak için bu stratejilerden yararlanabileceğini öne sürmektedirler. Sezgisel yöntemlerin bu kadar çeşitli alanlarda uygulanması, çok yönlülüklerinin ve karar verme sonuçlarını iyileştirme potansiyellerinin altını çizmektedir.

Sezgisel yöntemler, finans alanında yatırım kararlarında önemli bir rol oynamaktadır. Shah ve arkadaşları (2018), sezgisel süreçlerin doğasında var olan önyargıların, piyasa verimliliği algılarını nasıl etkileyerek optimal olmayan yatırım tercihlerine yol açabileceğini tartışmaktadır. Bu durum, Souza'nın (2023) sezgisel düşünceye dayanan psikolojik önyargıların yatırımcı performansını önemli ölçüde etkileyebileceğini gösteren araştırmasıyla da desteklenmektedir. Finansal bağlamlarda sezgisel ve bilişsel önyargılar arasındaki etkileşim, bu önyargıların karar verme üzerindeki olumsuz etkilerini azaltmak için farkındalık ve eğitim ihtiyacını göstermektedir.

Sezgisel yöntemlerin yapay zekâ sistemlerine entegrasyonu, zekânın ve karar vermenin doğası hakkında önemli soruları gündeme getirmektedir. Lee ve Wagenmakers (2014), ortaya çıkan sezgisel yöntem biliminin geleneksel rasyonellik kavramlarına meydan okuduğunu ve daha az bilginin bazen daha iyi karar sonuçlarına yol açabileceğini öne sürmektedirler. Bu bakış açısı, önyargılı sezgisel yöntemlerin belirli bağlamlarda çıkarımsal doğruluğu artırabileceğini öne süren Martignon ve Brighton'ın (2011) bulgularıyla uyumludur. Bu kavramların YZ çerçeveleri içinde araştırılması, yalnızca insan karar verme sürecini taklit etmekle kalmayıp aynı zamanda sezgisel uygulamanın nüanslarını anlayarak onu geliştiren sistemler geliştirmek için oldukça önemlidir.

YZ gelişmeye devam ettikçe, sezgisel karar verme ilkelerinin dahil edilmesi, karmaşık ortamlarda etkili bir şekilde gezinebilen sistemler oluşturmak için hayati önem taşıyacaktır. Lu ve Li'nin (2022) çalışması, karar vermenin insan zekâsına karşı test edildiği modern oyun, YZ sistemlerinde farklı teknik ve paradigmaları anlamının önemini vurgulamaktadır. Sezgisel karar vermenin altında yatan bilişsel süreçlere ilişkin devam eden bu araştırma, gerçek dünyadaki zorlukların üstesinden gelebilecek daha sofistike YZ sistemlerinin geliştirilmesini sağlayacaktır.

Sezgisel karar verme, çeşitli disiplinleri ve uygulamaları kapsayan çok yönlü bir çalışma alanıdır. Sezgisel yöntemlerin YZ sistemlerine entegrasyonu, karar verme verimliliğini ve etkinliğini artırmak için umut verici yollar sunmaktadır. Bununla birlikte bilişsel önyargılar ve hatalar için potansiyel, sezgisel yöntemlerin uygulandığı bağlamların dikkatli bir şekilde incelenmesini gerektirir. Araştırmalar gelişmeye devam ettikçe, insan bilişsel süreçleri ve YZ yetenekleri arasındaki etkileşim, çeşitli alanlarda karar vermenin geleceğini şekillendirecektir.

4. RASYONEL KARAR VERMEDE YAPAY ZEKÂ

Rasyonel karar verme modelleri uzun zamandan beri hem organizasyonel teorinin hem de yapay zekâ (YZ) gelişiminin temel taşlarından biri olmuştur. Bu modeller, karar vericilerin mevcut bilgi ve tercihlere dayanarak faydayı en üst düzeye çıkaracak şekilde hareket edecekleri varsayımına dayanmaktadır. YZ bağlamında, rasyonel karar verme paradigması çoğunlukla insan karar verme süreçlerini kopyalamayı veya geliştirmeyi amaçlayan algoritmalar aracılığıyla operasyonel hale getirilir. Bununla birlikte insanın bilişsel sınırlamaları ile algoritmik rasyonalite arasındaki etkileşim, karar verme bağlamlarında YZ'nin etkinliği ve etkileri hakkında önemli soruları ortaya çıkarmaktadır.

Karar vermede rasyonalite, “biçimsel” ve “özel rasyonalite” şeklinde iki temel biçime ayrılabilir (Baloğlu, 2011). Lindebaum ve arkadaşları (2020), biçimsel rasyonelliğin, algoritmaların iyi yürüttüğü kuralların ve prosedürlerin sistematik olarak uygulanmasıyla ilgili olduğunu, özel rasyonelliğin ise daha geniş bağlamı ve insan değerlerini dikkate alan karar vermenin niteliksel yönlerini içerdiğini ifade etmektedir. Bu ayrım, algoritmik karar alma süreçlerinin sınırlarını vurgulaması açısından çok önemlidir; bu süreçler verimlilik açısından üstün olabilir ancak genellikle insan karar alıcıların karmaşık durumlara getirdiği incelikli anlayıştan yoksundur. Biçimsel rasyonaliteye olan güven, nicel ölçütlere aşırı vurgu yapılmasına yol açarak, sağlam karar verme için gerekli olan nitel faktörleri potansiyel olarak bir kenara bırakabilir.

Yapay zekâ sistemlerinde, rasyonel karar verme modellerinin uygulanması genellikle önceden tanımlanmış parametrelere dayalı olarak beklenen faydayı en üst düzeye çıkarmayı amaçlayan matematiksel optimizasyon tekniklerinin kullanılmasını içerir. Zaffalon ve Miranda (2018), tercihlerin rasyonel karar vermedeki temel rolünü vurgulayarak, tercihlerin aksiyomatize edilmesinin hem ekonomi hem de YZ’de önemli bir başarı olduğunu belirtmektedirler. Bu teorik çerçeve, geçmiş verilere ve yerleşik tercihlere dayalı sonuçları tahmin ederek insan karar verme sürecini modellemeye çalışan birçok YZ algoritmasının temelini oluşturmaktadır. Bununla birlikte bu modellerin, genellikle bağlamdan ve öznel deneyimlerden etkilenen insan tercihlerinin karmaşıklığını doğru bir şekilde yansıtmasını sağlamakta zorluk yaşanmaktadır.

Rasyonel karar vermenin teorik temellerine rağmen, gerçek dünya uygulamaları sıklıkla beklenen ve gerçek sonuçlar arasındaki tutarsızlıkları ortaya koymaktadır. Örneğin Buçinca ve arkadaşlarının (2020) çalışması, vekil görevlerin ve öznel ölçümlerin YZ sistemlerinin yanıltıcı değerlendirmelerine yol açabileceğini vurgulayarak, insanlar ve YZ arasındaki etkileşimlerin tasarımının, karar verme etkinliğini önemli ölçüde etkilediğini öne sürmektedir. Bu bulgu, bilişsel önyargıların ve sezgisel yöntemlerin YZ sistemleri ile işbirliği içinde alınan kararların rasyonelliğini bozabileceği insan-YZ arayüzünü anlamanın önemini vurgulamaktadır. Bunların yanı sıra YZ’nin karar verme süreçlerine entegrasyonu, algoritmik tavsiyelere duyulan güven ve itimadın eleştirel bir şekilde incelenmesini de gerektirmektedir. Zhang ve arkadaşları (2020), YZ sistemleri tarafından sağlanan güven ve açıklamaların insan karar vericilerin doğruluğunu ve güven kalibrasyonunu nasıl etkilediğini araştırmışlardır. Söz konusu araştırma, karar vericilerin YZ’nin tavsiyelerini etkili bir şekilde kullanmak için YZ’nin yetenekleri ve sınırlamaları hakkında doğru bir zihinsel model geliştirmeleri gerektiğini göstermektedir. Bu dinamik, özellikle kararların sonuçlarının derin olabileceği sağlık ve ceza adaleti gibi yüksek riskli ortamlarla ilgilidir. Ancak buradaki zorluk, karar verme sonuçlarını geliştirmek için insan sezgisi ve algoritmik hassasiyetin bir arada var olabileceği dengeli bir ilişkiyi teşvik etmektir.

Örgütlerde YZ sistemlerinin uygulanması, rutin görevleri otomatikleştirerek ve veri odaklı içgörüler sağlayarak karar verme süreçlerini dönüştürebilir. Xu (2024), YZ’nin geniş veri kümelerini analiz ederek ve eğilimleri tahmin ederek karar verme sürecini geliştirdiğini ve böylece örgütlerin bilinçli stratejik seçimler yapmasını sağladığını savunmaktadır. Bununla birlikte, YZ’ye olan güven, özellikle algoritmaların kendisinde gömülü olabilecek önyargılarla ilgili sınırlamalarının anlaşılmasıyla hafifletilmelidir. Nishant ve arkadaşları (2023), algoritmik karar verme sürecinde hesap verebilirlik ve şeffaflık ihtiyacını vurgulayarak,

YZ'nin mevcut önyargıları devam ettirme potansiyelinin altını çizmektedirler. Bu endişe, YZ kararlarının etik sonuçlarının toplumsal sonuçları önemli ölçüde etkileyebileceği kamu örgütlerinde de görülebilmektedir (Sun ve Medaglia, 2019).

YZ'nin karar alma süreçlerine entegrasyonu, hesap verebilirliğin doğası hakkında da soruları gündeme getirmektedir. Gualdi ve Cordella (2021), YZ'nin resmi ve gayri resmi kuralları algoritmik senaryolara kodlayarak karar verme süreçlerini resmileştirdiğini ve bunun da kurumsal bağlamlarda hesap verebilirlik çizgilerini belirsizleştirebileceğini savunmaktadır. Bu değişim, YZ'ye giderek daha fazla bağımlı hale gelen ortamlarda hesap verebilirliğin nasıl tanımlandığının ve uygulandığının yeniden değerlendirilmesini gerektirmektedir. Buradaki zorluk, YZ sistemlerinin yalnızca etkili olmasını değil, aynı zamanda hizmet ettikleri paydaşlara karşı hesap verebilir olmasını sağlamakta yatmaktadır. Ayrıca, insan karar vericiler ve YZ sistemleri arasındaki ilişki karmaşık ve çok yönlüdür. Jiang ve arkadaşları (2023), YZ ve insan kararlarının alınma sırasının, çok aşamalı karar verme süreçlerinde prosedürel adalet algılarını nasıl etkileyebileceğini araştırmışlardır. Kararların sıralanmasının YZ tavsiyelerine olan güveni ve itimadı etkileyebileceğini öne sürmüşler ve insan-YZ işbirliğinin dinamiklerini anlamının önemine dikkat çekmişlerdir. Bu etkileşim, YZ'nin güven veya hesap verebilirliği engellemeden insan karar verme sürecini artırabileceği ortamları teşvik etmede kritik öneme sahip olduğunu göstermektedir.

Bu bağlamda rasyonel karar verme modeli hem organizasyon teorisinde hem de YZ gelişiminde temel bir kavram olmaya devam etmekte olduğu ifade edilebilir. Bununla birlikte, insan bilişinin karmaşıklığı, algoritmik rasyonelitenin sınırlamaları ve YZ karar vermenin etik sonuçları, bu modellerin pratikte nasıl işlediğine dair incelikli bir anlayış gerektirmekte olduğu söylenebilir. Bunun yanında YZ'nin psikoloji, örgütsel davranış ve benzer alanlarda yapılan araştırmalardan elde edilen bilgi/bulguları entegre etmesiyle; alan araştırmacılarına katkı sağlamak adına hem insan sezgisinin hem de algoritmik hassasiyetin güçlü yönlerinden yararlanan daha etkili karar verme çerçevelerinin geliştirilebileceği de dikkat çeken bir nokta olarak ifade edilebilir. Dolayısıyla var olan ve gelecekte ortaya çıkacak yeni/keşfedilmemiş konuların/temaların sürekli araştırılmasının da günümüz dünyasında karar vermenin geleceğini şekillendirmede önemli olacağı çıkarımında bulunulabilir.

5. SONUÇ

Yapay zekânın örgütlerdeki sezgisel ve rasyonel karar verme süreçlerine etkisi, örgütsel yapıları dönüştürmede kritik bir rol oynamaktadır. Yapay zekâ teknolojileri hem sezgisel hem de rasyonel karar mekanizma-

larını destekleyerek, belirsizlik ve zaman baskısı gibi karmaşık durumlarda etkili çözümler üretme potansiyeline sahiptir. Herbert Simon'ın sınırlı rasyonallite teorisi bağlamında, yapay zekânın sezgisel süreçleri güçlendirmesi, hızlı karar alma avantajını sunarken; bilişsel önyargıların azaltılmasına da katkı sağlayabilir. Bununla birlikte sezgisel karar verme süreçlerindeki hataların ve bilişsel önyargıların yapay zekâ tasarımında dikkatle ele alınması gerekmektedir. Diğer yandan, rasyonel karar verme perspektifinde, yapay zekâ sistemlerinin büyük veri analizi, karmaşık örüntü tanıma ve optimal çözümler üretme kapasitesi öne çıkmaktadır. Ancak bu sistemlerin biçimsel rasyonelliğe aşırı vurgu yapması, insani değerlerin göz ardı edilme riskini doğurabileceği için, insan-makine iş birliğinin dengeli bir şekilde tasarlanması ve etik çerçevelerin geliştirilmesi büyük önem taşımaktadır.

Gelecek perspektifinden bakıldığında, yapay zekânın örgütsel karar verme süreçlerindeki rolünün artarak devam etmesi beklenmektedir. Örgütler için hibrit karar verme sistemlerinin geliştirilmesi, insan sezgisi ile yapay zekânın analitik kapasitesini bir araya getirerek daha dengeli ve etkili bir yapı oluşturabilir. Bununla birlikte, hesap verebilirlik, şeffaflık ve algoritmalarındaki potansiyel önyargılar gibi zorlukların ele alınması, yapay zekâ sistemlerine duyulan güvenin kalibrasyonu açısından kritik öneme sahiptir. Yapay zekâ teknolojilerinin örgütsel kültür üzerindeki etkileri ve değişim yönetimi süreçlerine entegrasyonu, başarılı bir dönüşüm için temel unsurlar arasında yer almaktadır. Sonuç olarak, yapay zekâ destekli karar verme süreçlerinin etik ilkeler ve sürekli öğrenme mekanizmaları ile zenginleştirilmesi, gelecekte daha etkili ve sürdürülebilir organizasyonel karar alma yapılarını mümkün kılacaktır.

KAYNAKÇA

- Ahmad, M., Shah, S. Z. A. ve Abbass, Y. (2020). The Role of Heuristic-Driven Biases in Entrepreneurial Strategic Decision-Making: Evidence From an Emerging Economy. *Management Decision*, 59(3), 669-691. doi:10.1108/md-09-2019-1231
- Ahmad, M., Wu, Q., Naveed, M. ve Ali, S. (2022). Probing the Impact of Cognitive Heuristics on Strategic Decision-Making During the COVID-19 Pandemic: Evidence From an Emerging Economy. *International Journal of Social Economics*, 49(10), 1532-1550. doi:10.1108/ijse-10-2021-0636
- Albar, F. M. ve Jetter, A. (2009). Heuristics in Decision Making, 578-584. doi:10.1109/picmet.2009.5262123
- Bahreini, A. F., Cenfetelli, R. ve Cavusoglu, H. (2022). The Role of Heuristics in Information Security Decision Making. doi:10.24251/hicss.2022.587
- Balmaa, M. E. (2024). The Impact of Artificial Intelligence in Decision Making: A Comprehensive Review. *Epra International Journal of Economics Business and Management Studies*, 27-38. doi:10.36713/epra15747
- Baloğlu, D. F. (2011). Rasyonalite ve Ekonomik Sosyoloji. *Istanbul Journal of Sociological Studies*, (26), 217-226. <https://dergipark.org.tr/en/pub/iusoskon/issue/9524/119004> adresinden erişildi.
- Banks, A. P., Gamblin, D. ve Hutchinson, H. (2020). Training Fast and Frugal Heuristics in Military Decision Making. *Applied Cognitive Psychology*, 34(3), 699-709. doi:10.1002/acp.3658
- Brown, C., Nazeer, R., Gibbs, A., Pagé, P. ve Mitchell, A. R. (2023). Breaking Bias: The Role of Artificial Intelligence in Improving Clinical Decision-Making. *Cureus*. doi:10.7759/cureus.36415
- Buçınca, Z., Lin, P., Gajos, K. Z. ve Glassman, E. L. (2020). Proxy Tasks and Subjective Measures Can Be Misleading in Evaluating Explainable AI Systems. doi:10.1145/3377325.3377498
- Dietzmann, C. ve Duan, Y. (2022). Artificial Intelligence for Managerial Information Processing and Decision-Making in the Era of Information Overload. doi:10.24251/hicss.2022.720
- Gerlach, S., Hofmann, T., Fürweger, C. ve Schlaefer, A. (2022). AI-based Optimization for US-guided Radiation Therapy of the Prostate. *International Journal of Computer Assisted Radiology and Surgery*, 17(11), 2023-2032. doi:10.1007/s11548-022-02664-6
- Gualdi, F. ve Cordella, A. (2021). Artificial Intelligence and Decision-Making: The Question of Accountability. doi:10.24251/hicss.2021.281
- Guercini, S. (2022). Scope of Heuristics and Digitalization: The Case of Marketing Automation. *Mind & Society*, 21(2), 151-164. doi:10.1007/s11299-022-00291-x

- Guercini, S. (2023). Marketing Automation and the Scope of Marketers' Heuristics. *Management Decision*, 61(13), 295-320. doi:10.1108/md-07-2022-0909
- Hirshleifer, D., Levi, Y., Lourie, B. ve Teoh, S. H. (2019). Decision Fatigue and Heuristic Analyst Forecasts. *Journal of Financial Economics*, 133(1), 83-98. doi:10.1016/j.jfineco.2019.01.005
- Huy, V. Q. H. (2023). Understanding the Causes and Solutions of AI Induced Misinformation Impacting the Decision Making Behavior of Students. *Migration Letters*, 21(S1), 813-817. doi:10.59670/ml.v21is1.6452
- Jiang, L., Qin, X. S., Yam, K. C., Dong, X., Liao, W. ve Chen, C. (2023). Who Should Be First? How and When AI-human Order Influences Procedural Justice in a Multistage Decision-Making Process. *Plos One*, 18(7), e0284840. doi:10.1371/journal.pone.0284840
- Lee, M. ve Wagenmakers, E.-J. (2014). Heuristic Decision-Making, 224-236. doi:10.1017/cbo9781139087759.020
- Lindebaum, D., Vesa, M. ve Hond, F. d. (2020). Insights From “The Machine Stops” to Better Understand Rational Assumptions in Algorithmic Decision Making and Its Implications for Organizations. *Academy of Management Review*, 45(1), 247-263. doi:10.5465/amr.2018.0181
- Lu, Y. ve Li, W. (2022). Techniques and Paradigms in Modern Game AI Systems. *Algorithms*, 15(8), 282. doi:10.3390/a15080282
- Martignon, L. ve Brighton, H. (2011). Homo Heuristicus: Why Biased Minds Make Better Inferences, 2-26. doi:10.1093/acprof:oso/9780199744282.003.0001
- Nishant, R., Schneckenberg, D. ve Ravishankar, M. N. (2023). The Formal Rationality of Artificial Intelligence-Based Algorithms and the Problem of Bias. *Journal of Information Technology*, 39(1), 19-40. doi:10.1177/02683962231176842
- Oluwatamilore Popo – Olaniyan, Oluwafunmi Adijat Elufioye, Franciscamary Chinyere Okonkwo, Chioma Ann Udeh, Tobechukwu Francisca Eleogu, ve Funmilola Olatundun Olatoye. (2023). AI-Driven Talent Analytics For Strategic HR Decision-Making In The United States Of America: A Review. *International Journal of Management & Entrepreneurship Research*, 4(12), 607-622. doi:10.51594/ijmer.v4i12.674
- Riva, S., Monti, M. ve Antonietti, M. (2011). Simple Heuristics in Over-the-Counter Drug Choices: A New Hint for Medical Education and Practice. *Advances in Medical Education and Practice*, 59. doi:10.2147/amep.s13004
- Shah, S. Z. A., Ahmad, M. ve Mahmood, F. (2018). Heuristic Biases in Investment Decision-Making and Perceived Market Efficiency. *Qualitative Research in Financial Markets*, 10(1), 85-110. doi:10.1108/qrfm-04-2017-0033

- Shrestha, Y. R., Ben-Menahem, S. M. ve Krogh, G. v. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, 61(4), 66-83. doi:10.1177/0008125619862257
- Souza, V. V. (2023). Do Heuristics Affect Brazilian Investors' Decision-Making Process? *Review of Behavioral Finance*, 16(2), 282-299. doi:10.1108/rbf-09-2022-0210
- Sun, T. Q. ve Medaglia, R. (2019). Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence From Public Healthcare. *Government Information Quarterly*, 36(2), 368-383. doi:10.1016/j.giq.2018.09.008
- Volosova, A. V. ve Матюхина, Е. Н. (2020). Using Artificial Intelligence for Effective Decision-Making in Corporate Governance Under Conditions of Deep Uncertainty. *SHS Web of Conferences*, 89, 03008. doi:10.1051/shs-conf/20208903008
- Xu, S. (2024). Innovating Artificial Intelligence for Workforce Preparation and Knowledge Development. *Journal of Computer Science Research*, 6(2), 12-17. doi:10.30564/jcsr.v6i2.6663
- Zaffalon, M. ve Miranda, E. (2018). Desirability Foundations of Robust Rational Decision Making. *Synthese*, 198(S27), 6529-6570. doi:10.1007/s11229-018-02010-x
- Zhang, Y., Liao, Q. V. ve Bellamy, R. K. E. (2020). Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-assisted Decision Making. doi:10.1145/3351095.3372852
- Zimmermann, B., Shaw, D., Elger, B. S. ve Koné, I. (2021). The Use of Heuristics in Genetic Testing Decision-Making: A Qualitative Interview Study. *Plos One*, 16(11), e0260597. doi:10.1371/journal.pone.0260597

VERİ MADENCİLİĞİ ARAÇLARINA GENEL BİR BAKIŞ

Mahmut TOKMAK¹

1. GİRİŞ

Günlük olarak dünya genelinde üretilen veri miktarı sürekli olarak değişmekle birlikte, son yıllarda hızla artmaktadır. İnternet, mobil cihazlar, sensörler ve sosyal medya gibi teknolojik ilerlemeler, verinin artmasında ve depolama kapasitesinin artmasında etkili olmaktadır. Sayısal olarak bakıldığında bir dakikada dünya genelinde yaklaşık olarak 500 saatlik video, 500 milyon tweet, 4 milyon saatlik YouTube videosu ve 4.3 milyar internet kullanıcısının verileri üretilmektedir. Bu da günlük olarak yaklaşık olarak 2.5 quintillion (2.5×10^{18}) byte veri üretildiği anlamına gelmektedir (Domo 2024; Techjury 2023).

Bu hızlı büyüme, yalnızca verinin depolanması değil, aynı zamanda analiz edilerek anlamlı bilgilere dönüştürülmesi ihtiyacını doğurmuştur. Bu bağlamda, büyük veri (Big Data) ve veri madenciliği (Data Mining) kavramları, modern bilgi çağının en kritik bileşenleri haline gelmiştir. Geleneksel veri işleme yöntemleriyle yönetilemeyecek kadar büyük ve karmaşık olan bu veri setleri, yenilikçi teknolojilerin ve yaklaşımların geliştirilmesini zorunlu kılmıştır. Veri madenciliği, büyük veri setlerinden gizli desenler, eğilimler ve ilişkiler gibi anlamlı bilgilerin keşfedilmesini sağlayan bir analiz sürecidir. Bu süreç, istatistiksel yöntemler, makine öğrenimi algoritmaları ve yapay zeka tekniklerini kullanarak veriyi analiz eder. Veri madenciliğinin temel amacı, büyük ve karmaşık veri yığınlarından öngörüler elde ederek karar alma süreçlerini desteklemektir. Veri madenciliği, büyük veri analitiğiyle birlikte, iş dünyasından akademiye, sağlık sektöründen kamu hizmetlerine kadar pek çok alanda devrim yaratmaya devam etmektedir. Özellikle yapay zeka, makine öğrenimi ve kuantum bilişim gibi teknolojilerin entegrasyonu, büyük veri analitiğinde yeni bir çağ başlatmıştır (Sharmila ve Kanagaraj 2023; Tf ve Singh 2022).

Büyük veri çağında, karmaşık ve büyük veri setlerinden anlamlı bilgiler çıkarma yeteneği, endüstriler arasında başarının kritik bir unsuru haline gelmiştir. Açık kaynak veri madenciliği araçları, ileri düzey analitik

¹ Doç.Dr., Burdur Mehmet Akif Ersoy Üniversitesi, Bucak Zeliha Tolunay Uygulamalı Teknoloji ve İşletmecilik Yüksekokulu, Yönetim Bilişim Sistemleri, e-mahmuttokmak@mehmetakif.edu.tr, ORCID: 0000-0003-0632-4308.

yeteneklere erişimi kolaylaştırarak, bireylerin ve kuruluşların verilerden değer elde etmesini sağlamada önemli bir rol oynamaktadır. Bu araçların önemi yalnızca maliyet etkinliklerinde değil, aynı zamanda uyarlanabilirlik, ölçeklenebilirlik ve topluluk odaklı yeniliklerinde de yatmaktadır. Bu araçlar, veri analizi dünyasını dönüştürerek veri bilimciler, araştırmacılar ve işletmeler arasında yaratıcılığı, iş birliğini ve bilgi paylaşımını teşvik etmiştir (Berthold vd. 2008; H Witten vd. 2017; Prasad ve Prawal 2024; Ramadhan ve Niam 2023).

Açık kaynak veri madenciliği araçlarının sunduğu faydalar, sadece maliyet avantajıyla sınırlı değildir. Esnek ve modüler yapıları, kullanıcıların algoritmalarından veri işleme süreçlerine kadar her aşamada özelleştirmeler yapabilmesini mümkün kılar. Ayrıca, topluluk destekli projeler olarak sürekli geliştirilmeleri ve yeniliklere hızla uyum sağlamaları, bu araçların güncel kalmasını ve gelişen teknolojilere entegrasyonunu kolaylaştırmaktadır (Berthold vd. 2008).

Açık kaynak araçlarına olan ilginin artması, bu araçların başarıları, kullanım kolaylıkları ve uygulamaları hakkında daha derin bir anlayış gerekliliğini ortaya koymaktadır. Bu çalışma, en yaygın kullanılan açık kaynak veri madenciliği araçlarını incelemeyi, özelliklerini, avantajlarını ve gerçek dünya kullanım örneklerini değerlendirmeyi amaçlayarak, bu araçların veri analitiği ekosistemindeki dönüştürücü etkilerine ışık tutmayı hedeflemektedir. Bu bağlamda çalışmada kodlama bilgisi gerektirmeyen Weka ve Orange araçları ile kodlama bilgisi gerektiren R ve Scikit-learn araçları incelenmiştir. Daha öncesinde tamamen ücretsiz olup günümüzde bireysel kullanım için ücretsiz olup yetenekleri sınırlandırılmış versiyonları mevcut olan KNIME, Rapid Miner gibi birkaç araç çalışma kapsamı dışında tutulmuştur. Son olarak yapılan bu çalışma açık kaynak çözümlerinin sağladığı avantajların daha iyi anlaşılması, veri analitiği projelerinde doğru araç seçimi için rehber niteliğinde olacaktır.

2. LİTERATÜR ÖZETLERİ

Veri madenciliği alanı, yıllar içinde önemli gelişmelere sahne olmuş ve açık kaynak araçlarının geliştirilmesi ve uygulanmasına yönelik birçok çalışmaya konu olmuştur. Bu bölümde, açık kaynak veri madenciliği araçlarının evrimi, uygulamaları ve karşılaştırmalı analizleri üzerine yapılan temel çalışmalar ele alınmıştır.

Mierswa vd. (2006), RapidMiner'ın uçtan uca veri madenciliği sürecini nasıl kolaylaştırdığını vurgulamıştır. Çalışmada, RapidMiner'ın sürükle-bırak arayüzü ve gelişmiş makine öğrenimi algoritmalarını destekleyen yapısı sayesinde hem yeni başlayanlar hem de uzman kullanıcılar için uygun olduğunu belirtilmiştir.

Berthold vd. (2008), modüler ve ölçeklenebilir bir veri analitiği platformu olan KNIME'ı tanıtmıştır. Çalışmada, KNIME'in düğüm tabanlı arayüzünün iş akışı oluşturmayı nasıl kolaylaştırdığı ve kullanıcıların minimal kodlama ile karmaşık analizler yapmalarına olanak tanıdığı gösterilmiştir.

Demşar vd. (2013), veri görselleştirme ve analizi için tasarlanmış açık kaynak bir araç olan Orange'ın yeteneklerini incelemiştir. Çalışma, aracın eğitimdeki kullanımına odaklanarak, görsel programlama arayüzü ve etkileşimli görselleştirme özelliklerinin önemini vurgulamıştır.

Jovic vd. (2014), genel veri madenciliği görevleri için popüler olan altı ücretsiz yazılım aracı (RapidMiner, R, Weka, Orange, KNIME ve Scikit-Learn) üzerinde odaklanmışlardır. Araçları, algoritma desteği, kullanıcı dostu arayüzler ve topluluk desteği açısından değerlendirmişlerdir. RapidMiner'ı görsel arayüzüyle öne çıkarken gelişmiş algoritmalarda sınırlı bulmuşlardır. Weka'yı geniş algoritma desteğiyle güçlü, ancak kaynak gereksinimlerinin yüksek olarak nitelendirmişlerdir. R için istatistiksel analizde başarılı ancak öğrenmesi zor kanaatine varmışlardır. Orange'ın, görsel programlama ve Python entegrasyonu sunduğunu fakat widget sayısının sınırlı kaldığına vurgu yapmışlardır. KNIME aracının, genişletilebilirliğiyle öne çıktığını, Scikit-Learn'ün, hızlı ve esnek algoritmalar sunduğunu ancak programlama bilgisi gerektirdiğini ve sonuçta her aracın, farklı ihtiyaçlara göre tercih edilmesi gerektiğini belirtmişlerdir.

Al-Khoder ve Harmouch (2015), açık kaynak ve ücretsiz dört veri madenciliği aracını (R, RapidMiner, WEKA ve KNIME) platform, veri formatı, görselleştirme, popülerlik ve performans gibi kriterler üzerinden değerlendirmektedir. R, geniş giriş/çıkış formatları ve görselleştirme seçenekleriyle dikkat çekerken, WEKA en yüksek doğruluğu sağlayarak sınıflandırma performansında öne çıkmıştır. RapidMiner'ı, popülerlik açısından lider konumda olduğunu ve sürekli veri türlerini işlemek için ideal olarak nitelendirmişlerdir. KNIME, kullanıcı dostu bir yapı sunarak başlangıç seviyesindeki kullanıcılar için önerilmektedir. WEKA'nın, özellikle çoklu sınıf etiketleriyle çalışan veri kümelerinde üstün performans gösterdiği vurgulanmıştır. Çalışma, farklı veri madenciliği görevlerinde araç seçimi konusunda yol gösterici öneriler sunmaktadır.

Ranjan vd. (2017), veri madenciliği araçlarının kapsamlı bir analizini sunar ve RapidMiner, R, Knime, Orange ve Weka gibi beş açık kaynaklı araç üzerine odaklanır. Araçların güçlü ve zayıf yönleri, teknik özellikleri ve özel kullanımları değerlendirilmiştir. Çalışmada; RapidMiner, görsel programlama ve platform bağımsız çalışmasıyla öne çıkarken; R, geniş istatistiksel kütüphaneleri ve güçlü grafik özellikleriyle dikkat çeker. Weka, çok sayıda algoritma desteği sunar ancak klasik istatistikte zayıf kalır.

Orange, kolay öğrenilebilirliği ve görsel programlama desteğiyle kullanıcı dostudur. Knime, güçlü bir grafiksel kullanıcı arayüzü sunar ve özellikle yeni başlayanlar için önerilir. Çalışma, kullanıcıların farklı seviyelerdeki ihtiyaçlarına göre en uygun aracı seçmelerini kolaylaştırmayı amaçlamaktadır.

Bartschat vd. (2019), veri madenciliği araçlarının gelişimini, mevcut durumunu ve kullanıcı gruplarına göre sınıflandırılmasını ele almışlardır. Araçlar, iş zekası, akademik araştırma, algoritma geliştirme ve eğitim gibi çeşitli kullanıcı gruplarına hizmet edecek şekilde kategorize edilmiştir. KNIME, RapidMiner, TensorFlow ve Apache Spark gibi popüler araçlar arasında yer alırken, açık kaynak ve ticari araçlar arasındaki farklara da değinilmektedir. Ayrıca, veri madenciliği görevleri (örneğin sınıflandırma, kümeleme, ilişkilendirme) ve veri türlerine (örneğin grafikler, zaman serileri) göre farklı araçların nasıl kullanılabileceği tartışılmıştır. Büyük veri, IoT ve derin öğrenme gibi modern trendlerin veri madenciliği araçlarına etkileri vurgulanmıştır. Araştırmacılar ve uygulayıcılar için doğru aracı seçmenin, ihtiyaçlara uygun kategorilere göre yapılacak bir karşılaştırma ile kolaylaştırılabileceği belirtilmiştir.

Ratra ve Gulia (2020), veri madenciliği araçları WEKA ve Orange'ın sınıflandırma problemlerindeki performansını karşılaştırmalı olarak incelemektedir. Çalışmaya göre; WEKA, sınıflandırma ve model doğruluğu açısından Orange'dan daha iyi performans göstermiştir; özellikle Naïve Bayes (NB), Rastgele Orman (RO) ve K-En Yakın Komşu (KEYK) algoritmalarında öne çıkmıştır. Orange ise görsel programlama ve kullanıcı dostu arayüzüyle kolay kullanım sağlar, ancak daha sınırlı algoritma desteği ve veri görselleştirme özelliklerine sahiptir.

Tougui vd. (2020), altı farklı veri madenciliği aracı (Orange, Weka, RapidMiner, Knime, Matlab ve Scikit-Learn) ile altı makine öğrenimi tekniğini (Lojistik Regresyon, Destek Vektör Makineleri (DVM), KEYK, Yapay Sinir Ağları, NB ve RO) kullanarak kalp hastalıklarını sınıflandırmayı incelemişlerdir. Araştırmada, Matlab, en iyi performansı gösteren araç olarak Yapay sinir ağları modeliyle %85.86 doğruluk sağlamıştır. Knime, kullanıcı dostu arayüzü ve kolay kullanımı nedeniyle başlangıç seviyesindekilere önerilmiştir. Daha ileri düzey kullanıcılar için RapidMiner veya Knime önerilirken, programlama odaklı kullanıcılar için Matlab ve Scikit-Learn önerilmiştir.

Hosseini ve Sardo (2021), veri madenciliği ve makine öğrenimi araçlarının incelenmesi ve karşılaştırılması üzerine odaklanmaktadır. Araştırma, özellikle ağ saldırılarını tespit etmek için Knime ve Scikit-Learn araçlarını kullanarak bir durum çalışması içermektedir. WEKA, KNIME, Keel, Orange, Azure, IBM SPSS Modeler, R ve Scikit-Learn gibi araçların

özellikleri, destekledikleri algoritmalar ve kullanım kolaylıkları değerlendirilmiştir. Knime, görsel arayüz tercih edenler için; Scikit-Learn ise programlama becerisi yüksek kullanıcılar için önerilmiştir.

Dol ve Jawandhiya (2022), eğitim veri madenciliğinde kullanılan açık kaynak ve ticari veri madenciliği araçlarını karşılaştırmışlardır. Açık kaynak araçlar arasında Weka, KNIME, Orange ve RapidMiner gibi yazılımlar yer alırken, ticari araçlar arasında SPSS, SAS Enterprise Miner ve Oracle Data Mining bulunmaktadır. Araçlar, algoritma desteği, performans ölçütleri (doğruluk, hassasiyet, hatalar vb.), işletim sistemi uyumluluğu ve kullanıcı dostu özellikler açısından değerlendirilmiştir.

3. MATERYAL VE YÖNTEM

Seçilen veri madenciliği araçlarının performansı, her bir araç tarafından oluşturulan NB, KEYK, RO, DVM sınıflandırma modelleri kullanılarak değerlendirilmiştir. Bu bölümde kullanılan veri madenciliği araçları, kullanılan sınıflandırma yöntemleri, veri seti ve yöntemlerin performans metrikleri konu edilmiştir.

3.1. Veri Madenciliği Araçları

3.1.1. WEKA

Waikato Çevresi Bilgi Analizi (WEKA), Yeni Zelanda'daki Waikato Üniversitesi tarafından geliştirilen önde gelen bir açık kaynak yazılım paketidir. Veri madenciliği ve makine öğrenimi için tasarlanan WEKA, araştırmacılar, eğitimciler ve veri analizi alanında çalışan uygulayıcılar için sağlam ve kullanıcı dostu bir platform sunar. Veri ön işleme, sınıflandırma, regresyon, kümeleme, ilişkilendirme kuralları madenciliği ve görselleştirme gibi geniş bir yelpazede araçlar sunarak çeşitli analitik görevler için çok yönlü bir çözüm sağlar (H Witten vd. 2011).

WEKA'nın mimarisi, platformlar arası uyumluluk sağlayan ve diğer Java tabanlı uygulamalarla entegrasyonu kolaylaştıran Java programlama dili üzerine inşa edilmiştir. Yazılımın temel güçlü yönlerinden biri, kullanıcıların çeşitli veri işleme ve modelleme bileşenlerini zincirleyerek iş akışlarını özelleştirmelerine olanak tanıyan modüler tasarımıdır. WEKA, csv, arff, json gibi formatlarda veri girişini destekler ve bu veriler doğrudan analiz için yüklenebilir. Ayrıca, WEKA'nın Deney modülü, farklı makine öğrenimi algoritmalarının çeşitli parametre konfigürasyonları altında karşılaştırmalı değerlendirmesini sağlar, böylece yeniden üretilebilirlik ve kıyaslama imkânı sunar (H Witten vd. 2011).

WEKA'nın önemli bir avantajı, farklı uzmanlık seviyelerindeki kullanıcılara hitap edebilmesidir. Grafik kullanıcı arayüzü (GUI), karmaşık makine öğrenimi tekniklerini uygulamayı basitleştirirken, komut satırı arayüzü (CLI) ve Java API, daha fazla esneklik gerektiren ileri düzey kullanıcılar için uygundur. Araştırmacılar ve geliştiriciler, WEKA'nın çerçevesine kolayca özel algoritmalar entegre edebilir, bu da yenilik ve deney yapmayı teşvik eder (Hall vd. 2009).

WEKA, özellikle eğitim veri madenciliği, biyoinformatik ve öngörü analiz gibi akademik ve endüstriyel alanlarda geniş çapta benimsenmiştir. Popülarlığı, sürekli gelişimine katkıda bulunan güçlü bir kullanıcı ve geliştirici topluluğu tarafından da desteklenmektedir. Açık bilim ve metodolojik şeffaflığa vurgu yapan WEKA, kullanıcıların makine öğrenimi kavramlarını keşfetmelerine ve doğrulamalarına olanak tanıyan hem bir araştırma aracı hem de bir eğitim kaynağı olarak hizmet vermektedir (Bouckaert vd. 2010)

3.1.2. Orange

Orange, veri madenciliği ve makine öğrenimi için tasarlanmış açık kaynaklı, modüler ve kullanıcı dostu bir yazılım platformudur. Python tabanlı bir araç olan Orange, kullanıcıların veri analizi, görselleştirme ve modelleme süreçlerini kolaylıkla gerçekleştirmesine olanak tanır. Akademik ve endüstriyel uygulamalarda sıkça tercih edilen Orange, veri ön işleme, sınıflandırma, kümeleme, regresyon, ilişkilendirme kuralları madenciliği ve veri görselleştirme gibi geniş bir yelpazede araştırmaları destekler (Demšar vd. 2013)

Orange'ın en dikkat çekici özelliklerinden biri, görsel programlama yeteneğidir. Kullanıcılar, iş akışlarını grafiksel bir arayüzde sürükleyip bırak yöntemiyle oluşturabilir ve farklı veri madenciliği bileşenlerini bir araya getirebilir. Bu özellik, özellikle programlama bilgisi sınırlı olan kullanıcılar için güçlü bir çözüm sunar. Ayrıca, Python betiklerine tam erişim sağlayan Orange, ileri düzey kullanıcılar için esneklik ve genişletilebilirlik sağlar (Demšar vd. 2013).

Orange, bilimsel araştırmalardan öğretim ortamlarına kadar birçok alanda kullanılmaktadır. Araştırmacılar ve eğitimciler için güçlü bir görselleştirme desteği ve kapsamlı veri analizi araçları sunar. Topluluk tabanlı bir geliştirme yaklaşımına sahip olan Orange, sürekli güncellenen eklentileriyle kullanıcıların özelleştirilmiş ihtiyaçlarına cevap verir (Demšar vd. 2004, 2013). Orange, veri madenciliği ve makine öğrenimi süreçlerini kolaylaştıran, kullanıcı dostu ve esnek bir platform olarak öne çıkar. Akademik ve endüstriyel kullanımda sağladığı avantajlarla, veri odaklı problem çözme süreçlerinde önemli bir rol oynamaktadır.

3.1.3. R

R, istatistiksel hesaplama ve grafik oluşturma için geliştirilmiş açık kaynaklı bir programlama dilidir. Veri madenciliği ve makine öğrenimi alanlarında yaygın olarak kullanılan R, kullanıcılarına zengin bir paket ekosistemi ve güçlü analiz araçları sunar. Veri ön işleme, kümeleme, sınıflandırma, regresyon, ilişkilendirme kuralları madenciliği ve zaman serisi analizi gibi geniş bir yelpazede veri madenciliği görevlerini destekleyen R, akademik ve endüstriyel kullanımlarda etkili bir çözümdür (Ihaka ve Gentleman 1996).

R'nin veri madenciliğindeki başarısının temelinde esneklik ve genişletilebilirlik yer alır. CRAN (Comprehensive R Archive Network) üzerinde yer alan ggplot2, dplyr, tidyr, caret, rpart, e1071 gibi bir çok paketler, makine öğrenimi ve veri madenciliği süreçlerinde kritik öneme sahiptir. Bu paketler, kullanıcıların karmaşık algoritmaları kolaylıkla uygulamasını sağlar ve aynı zamanda yüksek düzeyde özelleştirme ve görselleştirme imkânı sunar (Kuhn 2008; Wickham 2016).

R, güçlü istatistiksel altyapısı nedeniyle akademik araştırmalarda sıklıkla tercih edilmektedir. Bunun yanı sıra, finans, biyoinformatik, sağlık analitiği ve pazarlama gibi endüstriyel alanlarda da yaygın olarak kullanılmaktadır. Programlama dili olmaktan öte, R aynı zamanda açık kaynaklı bir topluluğun parçası olarak kullanıcılar arasında bilgi paylaşımını ve işbirliğini teşvik etmektedir (Chang 2018; Wickham 2016).

3.1.4. Scikit-learn

Scikit-learn, Python programlama dili üzerinde çalışan açık kaynaklı bir makine öğrenimi kütüphanesidir. Veri madenciliği, veri analitiği ve makine öğrenimi süreçlerinde geniş bir kullanım alanına sahiptir. Sınıflandırma, kümeleme, regresyon, boyut indirgeme ve model değerlendirme gibi temel veri madenciliği görevlerini destekleyen Scikit-learn, basit ve tutarlı bir API tasarımı ile kullanıcı dostu bir platform sunar (Bisong 2019; Pedregosa vd. 2011)

Scikit-learn, NumPy, SciPy ve matplotlib gibi güçlü Python kütüphanelerinin üzerine inşa edilmiştir. Bu altyapı, Scikit-learn'ün yüksek performanslı hesaplamalar ve etkili veri görselleştirme sağlamasına olanak tanır. Ayrıca, kullanıcıların farklı makine öğrenimi algoritmalarını kolayca uygulayabilmesine ve karşılaştırabilmesine imkân tanır. Genişletilebilir yapısı sayesinde araştırmacılar ve geliştiriciler, Scikit-learn'ü özelleştirilmiş ihtiyaçlar için adapte edebilir (Bisong 2019; Kramer 2016).

Scikit-learn, akademik ve endüstriyel uygulamalarda yaygın olarak kullanılmaktadır. Akademik dünyada, veri madenciliği ve makine öğreni-

mi eğitimi için ideal bir araçtır. Endüstride ise finans, sağlık, pazarlama ve doğal dil işleme gibi birçok alanda veri odaklı karar verme süreçlerini destekler. Kütüphane, aktif bir topluluk tarafından desteklenmekte ve sürekli güncellenerek yeni algoritmalar ve özelliklerle zenginleştirilmektedir (Bisong 2019; Kramer 2016; Pedregosa vd. 2011).

3.2. Kullanılan Yöntemler

3.2.1. Naive Bayes

Naive Bayes sınıflandırıcı, olasılıksal öğrenme temelli bir makine öğrenmesi algoritmasıdır ve Bayes teoremine dayanır. Algoritma, sınıflandırma işlemi sırasında tüm özelliklerin birbirinden bağımsız olduğunu varsayar; bu nedenle “naive” olarak adlandırılır. Özelliklerin bağımsız olduğu varsayımı her zaman geçerli olmasa da, algoritma genellikle oldukça etkili sonuçlar verir. Naive Bayes, özellikle metin sınıflandırması (örneğin, spam tespiti) ve duygu analizi gibi yüksek boyutlu veri kümelerinde başarılıdır. Algoritma, hem kategorik hem de sürekli veri ile çalışabilir ve sınıf önceliklerini ve özelliklerin sınıfa özgü olasılıklarını değerlendirerek hızlı ve basit bir sınıflandırma sağlar. Özellikle küçük veri kümelerinde bile etkili çalışması ve hızlı hesaplama avantajı, Naive Bayes’i yaygın kullanılan bir yöntem haline getirir (Rish 2001).

3.2.2. K-En Yakın Komşu Algoritması

K-En Yakın Komşu sınıflandırıcı, örneklerin sınıflarını tahmin etmek için veri noktalarının en yakın komşularına dayanan, örnek tabanlı bir makine öğrenmesi algoritmasıdır. Algoritma, yeni bir veri noktası geldiğinde, bu noktaya en yakın k komşuyu belirler ve çoğunluk sınıfına göre sınıflandırma yapar. Uzaklık ölçümü için genellikle Öklid uzaklığı kullanılır, ancak Minkowski veya Manhattan uzaklıkları gibi farklı metrikler de uygulanabilir. KEYK, eğitim sırasında herhangi bir parametre öğrenmediği için parametrik olmayan bir algoritmadır ve “lazy learner” olarak adlandırılır. Özellikle düşük boyutlu ve küçük veri kümelerinde etkili olan algoritma, büyük veri setlerinde yüksek hesaplama maliyeti nedeniyle yavaşlayabilir. K değeri, algoritmanın performansını etkileyen kritik bir parametredir; küçük k değerleri modelin aşırı öğrenmesine (overfitting), büyük k değerleri ise düşük genelleştirme yeteneğine neden olabilir (Altman 1992).

3.2.3. Rastgele Orman Algoritması

Rastgele Orman, birden çok karar ağacını bir araya getirerek çalışan, topluluk öğrenme (ensemble learning) tabanlı güçlü bir sınıflandırma algoritmasıdır. Algoritma, her bir ağacın farklı bir veri alt kümesi ve özellik alt kümesi ile eğitilmesi için bootstrap (bagging) yöntemi kullanır. Bu çeşitlilik, modelin genelleme yeteneğini artırır ve aşırı öğrenme (overfitting) riskini azaltır. Her bir karar ağacı, sınıflandırma sırasında bağımsız bir tahmin yapar ve RO, tüm ağaçların oy çokluğuna (majority voting) dayalı olarak nihai sınıf tahminini yapar. Hem sınıflandırma hem de regresyon problemleri için kullanılabilen algoritma, yüksek boyutlu veriler ve eksik verilerle etkili bir şekilde çalışabilir. Özellikle karar ağaçlarının hassas olduğu veri varyansını ve gürültüyü azaltarak daha kararlı sonuçlar üretir. RO, hesaplama maliyetinin artırılmasını göze alarak, esnekliği ve doğruluğu nedeniyle makine öğrenmesinde yaygın olarak kullanılan bir yöntemdir (Breiman 2001).

3.2.4. Destek Vektör Makineleri

Destek Vektör Makineleri, sınıflar arasında en iyi ayrımı sağlayan bir karar sınırı oluşturarak çalışan güçlü bir sınıflandırma algoritmasıdır. DVM, sınıflar arasındaki marj adı verilen boşluğu maksimize eden bir hiper düzlem bulmayı amaçlar. Eğer veri doğrusal olarak ayrılabilir değilse, algoritma, veriyi daha yüksek boyutlu bir uzaya projelendirmek için çekirdek (kernel) fonksiyonlarını kullanır. Yaygın çekirdek türleri arasında doğrusal (linear), radyal bazlı (RBF), polinomial ve sigmoid çekirdekler bulunur. DVM, yüksek boyutlu veri kümelerinde etkili çalışabilir ve aşırı öğrenmeye (overfitting) karşı oldukça dirençlidir. Ancak, büyük veri setlerinde hesaplama maliyetleri artabilir ve hiperparametre ayarı (C ve gamma) hassas bir şekilde yapılmalıdır. DVM, genellikle sınıflandırma problemleri (örneğin, metin sınıflandırması veya görüntü tanıma) için tercih edilen bir yöntemdir (Cortes ve Vapnik 1995).

3.3. Veri Seti

Çalışmada Breast Cancer Wisconsin (BSW) veri seti kullanılmıştır. BCW veri seti, meme kanseri teşhisinde kullanılan tıbbi bir veri setidir ve makine öğrenmesi çalışmaları için sıklıkla tercih edilmektedir. Veri seti, hücre çekirdeğinin boyut, şekil, doku, çevre, alan ve pürüzlülük gibi özelliklerini içeren 30 adet sayısal özelliğe sahiptir. Bu özellikler, bir hastanın biyopsi sonuçlarına dayalı olarak elde edilmiş ve hastalığın iyi huylu (Benign) veya kötü huylu (Malignant) olarak sınıflandırılmasına olanak tanımaktadır. Veri seti, sınıflandırma algoritmalarının karşılaştırılması, tanı için en iyi özelliklerin belirlenmesi ve tıbbi teşhis sistemlerinin geliştiril-

mesi için güçlü bir temel sağlar. Dr. William H. Wolberg tarafından Wisconsin-Madison Üniversitesi'nde toplanmış olan veri seti, UCI Machine Learning Repository'de yayınlanmıştır ve makine öğrenmesi topluluğu için önemli bir kaynak olarak değerlendirilmektedir (Wolberg vd. 1993).

3.4. Performans Değerlendirme

Çalışmada kullanılan Performans değerlendirme metrikleri aşağıda açıklanmıştır:

- Doğruluk (Accuracy): Modelde doğru tahmin edilen alanların toplam veri kümesine oranı ile hesaplanmaktadır. Denklem 1'de doğruluk formülü gösterilmektedir.

$$\text{Doğruluk} = \frac{DP + DN}{DP + YN + YP + DN} \quad (1)$$

Kesinlik (Precision): Pozitif tahminlerin doğruluğu Kesinlik formülü ile bulunur. Denklem 2'de kesinlik formülü gösterilmektedir.

$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (2)$$

Duyarlılık (Recall): Model tarafından doğru şekilde tespit edilen pozitif örneklerin oranına denir. Denklem 3'te duyarlılık formülü gösterilmektedir.

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (3)$$

F1-Skor (F1-Score): F1 skor değeri hesaplanırken kesinlik ve duyarlılık değerleri bir arada kullanılarak yeni bir değer hesaplanmaktadır. Denklem 4'te F1-Skor formülü gösterilmektedir.

$$\text{F1 - Skor} = 2 \times \frac{\text{kesinlik} \times \text{duyarlılık}}{\text{kesinlik} + \text{duyarlılık}}$$

4. SONUÇLAR VE TARTIŞMA

Çalışmada NB, KEYK, RO ve DVM algoritmaları kullanılmıştır. Algoritmalar koşturulurken 10 fold cross validation kullanılmış ve parametre ayarlaması yapılmamıştır. Yani default parametreler kullanılmıştır. Elde edilen test metrikleri Tablo 1'de verilmiştir.

Tablo 1. Performans Metrikleri

Araçlar	Yöntemler	Kesinlik	Duyarlılık	F-Skoru	Doğruluk
Weka	NB	0,930	0,930	0,930	0,929
	KEYK	0,96	0,96	0,96	0,959
	RO	0,965	0,965	0,965	0,965
	DVM	0,977	0,977	0,977	0,977
Orange	NB	0,940	0,939	0,940	0,939
	KEYK	0,948	0,943	0,942	0,943
	RO	0,996	0,996	0,996	0,992
	DVM	0,990	0,990	0,990	0,992
Sci-kit-learn	NB	0,940	0,931	0,933	0,938
	KEYK	0,924	0,915	0,917	0,924
	RO	0,962	0,957	0,959	0,963
	DVM	0,934	0,898	0,910	0,919
R	NB	0,946	0,952	0,949	0,935
	KEYK	0,926	0,961	0,943	0,927
	RO	0,960	0,977	0,968	0,959
	DVM	0,975	0,98	0,978	0,971

Tablo 1’de elde edilen sonuçlara bakıldığında; Weka, özellikle KEYK ve DVM algoritmalarında güçlü sonuçlar elde etmiştir. RO algoritmasında iyi sonuçlar sağlasa da Orange’a göre geride kalmıştır. Orange, özellikle RO ve DVM algoritmalarında diğer araçlara kıyasla en iyi sonuçları sağlamıştır. Python, RO algoritmasında güçlü olsa da, diğer algoritmalarda daha düşük performans göstermiştir. R, özellikle NB, DVM ve RO algoritmalarında güçlü sonuçlar sağlamıştır.

Aynı algoritmaları kullanan araçların birbirine yakın olsa da farklı sonuçlar elde etmeleri, araçların “default” parametre seçimleriyle ilgili olduğu kanaatine varılmıştır.

Veri madenciliği ve makine öğrenimi araçları, kullanıcı hedeflerine ve deneyim seviyelerine göre farklı özellikler sunmaktadır. R Studio ve Scikit-learn gibi kod tabanlı araçlar, esneklik ve geniş algoritma desteği sunarken; Orange ve WEKA gibi görsel araçlar kullanıcı dostu arayüzleri ile daha hızlı öğrenim ve uygulama imkânı sağlamaktadır.

Çalışmaya konu olan araçlardan Weka ve Orange ile temel veri madenciliği bilgisine sahip bir kullanıcı tarafından veri madenciliği yapılabileceği görülmüştür. Özellikle Orange programı sürükle bırak mantığı ile çalıştığı ve kullanıcı dostu bir arayüze sahip olduğundan Wekaya göre

daha kullanışlı bulunmuştur. R ve Scikit-learn araçları ise veri madenciliği bilgisine sahip olmanın yanısıra kodlama bilgisi gerektirmektedir.

Sonuç olarak; Orange ve WEKA, görselleştirme odaklı hızlı analizler yapmak isteyenler ve öğrenme aşamasındaki kullanıcılar için idealdir. R ve Scikit-learn, daha ileri düzey kullanıcılar ve esneklik arayanlar için uygundur. Dolayısıyla her araç, kullanıcı ihtiyaçlarına ve deneyim seviyelerine göre belirli avantajlar sunar. Bu nedenle, doğru aracı seçmek kullanıcı hedeflerine bağlıdır.

KAYNAKÇA

- Al-Khoder, Ahmad, ve Hazar Harmouch. 2015. "Evaluating four of the most popular open source and free data mining tools". *International Journal of Academic Scientific Research* 3(10):13-23.
- Altman, N. S. 1992. "An Introduction to Kernel and Nearest-Neighbor Non-parametric Regression". *The American Statistician* 46(3):175-85. doi: 10.1080/00031305.1992.10475879.
- Bartschat, Andreas, Markus Reischl, ve Ralf Mikut. 2019. "Data Mining Tools". *WIRES Data Mining and Knowledge Discovery* 9(4):e1309. doi: 10.1002/widm.1309.
- Berthold, Michael R., Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinl, Peter Ohl, Christoph Sieb, Kilian Thiel, ve Bernd Wiswedel. 2008. "KNIME: The Konstanz Information Miner". Ss. 319-26 içinde *Data Analysis, Machine Learning and Applications*, editör C. Preissach, H. Burkhardt, L. Schmidt-Thieme, ve R. Decker. Berlin, Heidelberg: Springer.
- Bisong, Ekaba. 2019. "Introduction to Scikit-Learn". Ss. 215-29 içinde *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, editör E. Bisong. Berkeley, CA: Apress.
- Bouckaert, Remco R., Eibe Frank, Mark A. Hall, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, ve Ian H. Witten. 2010. "WEKA—Experiences with a Java Open-Source Project". *The Journal of Machine Learning Research* 11:2533-41.
- Breiman, Leo. 2001. "Random forests". *Machine learning* 45(1):5-32.
- Chang, Winston. 2018. *R graphics cookbook: practical recipes for visualizing data*. O'Reilly Media.
- Cortes, Corinna, ve Vladimir Vapnik. 1995. "Support-Vector Networks". *Machine Learning* 20(3):273-97. doi: 10.1007/BF00994018.
- Demšar, Janez, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinović, Martin Možina, Matija Polajnar, Marko Toplak, ve Anže Starič. 2013. "Orange: data mining toolbox in Python". *the Journal of machine Learning research* 14(1):2349-53.
- Demšar, Janez, Blaž Zupan, Gregor Leban, ve Tomaz Curk. 2004. "Orange: From Experimental Machine Learning to Interactive Data Mining". Ss. 537-39 içinde *Knowledge Discovery in Databases: PKDD 2004*. C. 3202, *Lecture Notes in Computer Science*, editör J.-F. Boulicaut, F. Esposito, F. Giannotti, ve D. Pedreschi. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dol, Sunita M., ve Pradeep M. Jawandhiya. 2022. "Use of data mining tools in educational data mining". Ss. 380-87 içinde *2022 Fifth International Con-*

- ference on Computational Intelligence and Communication Technologies (CCICT)*. IEEE.
- Domo. 2024. “Domo”. *Data Never Sleeps 8.0*. Geliş tarihi 24 Nisan 2023 (<https://www.domo.com/learn/data-never-sleeps-8>).
- H Witten, Ian, Eibe Frank, Mark A Hall, ve Christopher J Pal. 2011. *Data mining practical machine learning tools and techniques*. Morgan Kaufmann.
- H Witten, Ian, Eibe Frank, Mark A Hall, ve Christopher J Pal. 2017. “Data mining practical machine learning tools and techniques”.
- Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, ve Ian H. Witten. 2009. “The WEKA data mining software: an update”. *ACM SIGKDD explorations newsletter* 11(1):10-18.
- Hosseini, Soodeh, ve Saman Rafiee Sardo. 2021. “Data Mining Tools -a Case Study for Network Intrusion Detection”. *Multimedia Tools and Applications* 80(4):4999-5019. doi: 10.1007/s11042-020-09916-0.
- Ihaka, Ross, ve Robert Gentleman. 1996. “R: A Language for Data Analysis and Graphics”. *Journal of Computational and Graphical Statistics* 5(3):299-314. doi: 10.1080/10618600.1996.10474713.
- Jovic, Alan, Karla Brkic, ve Nikola Bogunovic. 2014. “An overview of free software tools for general data mining”. Ss. 1112-17 içinde *2014 37th International convention on information and communication technology, electronics and microelectronics (MIPRO)*. IEEE.
- Kramer, Oliver. 2016. “Scikit-Learn”. Ss. 45-53 içinde *Machine Learning for Evolution Strategies*, editör O. Kramer. Cham: Springer International Publishing.
- Kuhn, Max. 2008. “Building predictive models in R using the caret package”. *Journal of statistical software* 28:1-26.
- Mierswa, Ingo, Michael Wurst, Ralf Klinkenberg, Martin Scholz, ve Timm Euler. 2006. “YALE: Rapid Prototyping for Complex Data Mining Tasks”. Ss. 935-40 içinde *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. Philadelphia PA USA: ACM.
- Owen, Sean, Robin Anil, Ted Dunning, Ellen Friedman, ve MAHOUT IN ACTION. 2011. “Manning Publications”.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, ve Vincent Dubourg. 2011. “Scikit-learn: Machine learning in Python”. *the Journal of machine Learning research* 12:2825-30.
- Prasad, Smrity, ve Kashvi Prawal. 2024. “Big Data Analytics Lifecycle”. Ss. 100-118 içinde *Big Data Analytics Techniques for Market Intelligence*. IGI Global.

- Ramadhan, Gildan Jaya Muhammad, ve Sukron Niam. 2023. “Big Data Analytics: Techniques, Tools, and Applications in Various Industries”. *Jurnal Ar Ro'is Mandalika (Armada)* 3(2):56-65.
- Ranjan, Rohit, Swati Agarwal, ve S. Venkatesan. 2017. “Detailed analysis of data mining tools”. *Int. J. Eng. Res. Technol* 6(5).
- Ratra, Ritu, ve Preeti Gulia. 2020. “Experimental evaluation of open source data mining tools (WEKA and Orange)”. *International Journal of Engineering Trends and Technology* 68(8):30-35.
- Rish, Irina. 2001. “An empirical study of the naive Bayes classifier”. Ss. 41-46 içinde *IJCAI 2001 workshop on empirical methods in artificial intelligence*. C. 3. Seattle, WA, USA;
- Sharmila, S., ve A. Kanagaraj. 2023. “Big Data in Data Mining Techniques—A Survey”. *Kristu Jayanti Journal of Computational Sciences (KJCS)* 68-80.
- Techjury. 2023. “Techjury”. *Big Data Statistics*. Geliş tarihi 06 Nisan 2023 (<https://techjury.net/blog/big-data-statistics/>).
- Tf, Michael Raj, ve Yaduvir Singh. 2022. “An Exploration on Big Data Analysis and Data Mining Methods”. Ss. 1-6 içinde *2022 International Conference on Futuristic Technologies (INCOFT)*. IEEE.
- Tougui, Ilias, Abdelilah Jilbab, ve Jamal El Mhamdi. 2020. “Heart Disease Classification Using Data Mining Tools and Machine Learning Techniques”. *Health and Technology* 10(5):1137-44. doi: 10.1007/s12553-020-00438-1.
- Wickham, Hadley. 2016. “Data Analysis”. Ss. 189-201 içinde *ggplot2, Use R!* Cham: Springer International Publishing.
- Wolberg, William, Olvi Mangasarian, Nick Street, ve W. Street. 1993. “Breast Cancer Wisconsin (Diagnostic)”.

BÜYÜK VERİ, VERİ ANALİTİĞİ VE İŞLETMELER AÇISINDAN ÖNEMİ

Levent KOŞAN¹

1. GİRİŞ

Büyük veri ve veri analitiği, günümüz dijital çağında işletmelerin başarısı için kritik öneme sahip teknolojilerdir. Dijitalleşmenin hızla ilerlemesiyle birlikte, verinin üretimi ve depolanması büyük bir ivme kazanmış, veriye dayalı karar alma süreçleri işletmelerin stratejik yönelimlerini belirlemiştir. Özellikle internetin ve mobil cihazların yaygınlaşması, işletmelerin topladığı veri hacmini artırmış ve veri, işletmelerin en değerli varlıklarından biri haline gelmiştir. Bu bağlamda, büyük veri ve veri analitiği, sadece verilerin toplanması ve depolanmasının yanında bu verilerin anlamlı bilgiler haline getirilmesi ve işletme stratejilerine entegre edilmesi sürecini kapsamaktadır (Mayer-Schönberger & Cukier, 2013).

Büyük veri, genellikle yüksek hacim, hız ve çeşitlilikteki verileri tanımlar ve bu veriler, yapısal, yarı yapısal ve yapılandırılmamış olabilir. Bu veri setleri, doğru analiz yöntemleri kullanılarak anlamlı bilgilere dönüştürülmekte ve işletmelere çeşitli stratejik avantajlar sağlamaktadır. Veri analitiği ise bu büyük veri kümelerinin işlenmesi, analiz edilmesi ve işletmelere faydalı çıkarımlar sunulması sürecidir. Veri analitiği, işletmelere, veriye dayalı kararlar alma, müşteri davranışlarını tahmin etme, operasyonel verimliliği artırma, pazarlama stratejilerini optimize etme gibi birçok farklı alanda katkı sağlamaktadır (Davenport & Harris, 2007).

Büyük veri, genellikle üç temel özellik ile tanımlanır: hacim (volume), hız (velocity) ve çeşitlilik (variety). Hacim, veri miktarını ifade ederken, hız, verinin işlenme hızını ve çeşitlilik ise verinin kaynaklarını ve biçimlerini belirtir (Gartner, 2012). Bu üç boyut, büyük veri analizinin neden bu kadar karmaşık ve zorlu olduğunu açıklamaktadır. Büyük verinin bu özellikleri, hem veri toplama hem de veri analizi süreçlerinde işletmelere önemli teknik zorluklar yaratırken, aynı zamanda veriden elde edilen sonuçların değerini artırmaktadır.

¹ Prof. Dr., Mersin Üniversitesi, Turizm Fakültesi, Turizm İşletmeciliği Bölümü, lko-san@mersin.edu.tr, ORCID: 0000-0001-7630-6354

Veri analitiği, veriden anlamlı bilgi çıkarma sürecidir ve farklı aşamalara ayrılan analiz süreçlerini kapsar. Bu analizler, işletmelerin mevcut performanslarını değerlendirmelerini, gelecekteki eğilimleri tahmin etmelerini ve stratejik kararlarını daha veriye dayalı bir şekilde almalarını sağlar.

Büyük veri, işletmelere yalnızca operasyonel verimlilik sağlamaz; aynı zamanda müşteri deneyimini iyileştirme, yeni iş fırsatları yaratma ve stratejik kararları daha etkili bir şekilde alma imkanı tanır. Veriyi doğru bir şekilde kullanabilen işletmeler, rekabet avantajı elde edebilir ve pazar paylarını artırabilir. İşletmeler, büyük veri analitiği ile müşteri tercihlerini daha iyi anlayabilir, hedef odaklı pazarlama stratejileri geliştirebilir ve tedarik zincirlerini optimize edebilir (McAfee & Brynjolfsson, 2012).

Büyük veri analitiği, birçok farklı sektörde uygulama alanı bulmaktadır. Perakende sektörü, finans, sağlık, enerji ve kamu sektörü, büyük verinin potansiyelinden yararlanan başlıca alanlardır. Perakende sektöründe, müşteri verilerinin analiz edilmesi, envanter yönetimi ve fiyat optimizasyonu gibi konularda önemli faydalar sağlamaktadır. Finans sektöründe ise risk yönetimi ve portföy optimizasyonu gibi alanlarda büyük veri analitiği önemli bir rol oynamaktadır. Sağlık sektöründe büyük veri, hastaların tıbbi geçmişlerini ve tedavi sonuçlarını analiz ederek, daha doğru teşhisler ve tedavi planları oluşturulmasına yardımcı olabilir. Ayrıca, enerji sektöründe, akıllı sayaçlar ve sensörlerden toplanan veriler, enerji tüketimindeki anormallikleri tespit edebilir ve enerji verimliliğini artırmak için stratejiler geliştirebilir. Kamu sektöründe ise, trafik yönetimi, suç öncesi tahminleme ve diğer sosyal hizmetlerin iyileştirilmesi gibi alanlarda büyük verinin potansiyelinin geniş olduğu görülmektedir. (Chen et al., 2012).

Büyük verinin işletmelere sunduğu fırsatlarla birlikte, bu verinin işlenmesi ve analiz edilmesi sırasında karşılaşılan zorluklar da büyüktür. Veri güvenliği, veri kalitesi, depolama, entegrasyon ve analiz gibi konular, büyük veri projelerinin başarısını etkileyen önemli faktörlerdir. Örneğin, veri güvenliği ve gizliliği, özellikle kişisel verilerin toplandığı ve işlendiği sektörlerde büyük bir endişe kaynağıdır. Bununla birlikte, veri entegrasyonu da büyük veri projelerinde sıklıkla karşılaşılan bir diğer zorluktur; farklı veri kaynaklarının birleşmesi ve uyumlu hale getirilmesi gerekmektedir. Ayrıca, büyük verinin analizi için gerekli olan ileri düzeydeki teknolojik altyapılar ve yetkinlikler de işletmeler için önemli bir maliyet unsuru oluşturabilir. Bu nedenle, büyük veri projelerinin başarılı olabilmesi için işletmelerin yalnızca teknolojiye değil, aynı zamanda insan kaynağına ve organizasyonel yapıya yatırım yapmaları gerekmektedir (McKinsey & Company, 2011).

Bu bölümün amacı, büyük veri ve veri analitiği teknolojilerinin işletmelere sunduğu fırsatları ve karşılaşılan zorlukları incelemektir. Ayrıca,

büyük verinin işletmeler için nasıl stratejik bir kaynak haline geldiği, bu veriden nasıl değer elde edilebileceği ve hangi alanlarda uygulanabileceği üzerinde durulacaktır.

2. BÜYÜK VERİ KAVRAMI

Büyük veri, temel olarak üç V (Volume, Velocity, Variety) ile tanımlanır: hacim, çeşit ve hız. Bu üç temel özellik, büyük veriyi diğer veri türlerinden ayıran önemli faktörlerdir.

2.1. Hacim (Volume)

Büyük verinin temel özelliklerinden biri olan **hacim**, veri kümesinin boyutunu ifade eder ve bu, büyük veriyi anlamada en kritik bileşendir. Günümüzde işletmeler, dijitalleşme ve IoT (Internet of Things) gibi gelişen teknolojiler sayesinde her geçen gün büyük miktarda veri üretmektedir. Bu veri hacminin artışı, veri toplama, saklama, işleme ve analiz etme gereksinimlerini de büyük ölçüde etkilemektedir. Veri hacmi, özellikle internet kullanımının artması, sosyal medya etkileşimleri, mobil cihazlardan gelen veriler, video içerikleri ve IoT cihazlarının üretimiyle hızla artmaktadır. Sosyal medya platformlarında, her saniye milyonlarca kullanıcı etkileşimi (paylaşımlar, beğeniler, yorumlar) gerçekleşmekte ve bu etkileşimler hem metin hem de görsel veriler olarak çok büyük hacimler oluşturmaktadır (Mayer-Schönberger & Cukier, 2013). Bu veriler, sadece kullanıcı davranışlarını izlemekle kalmaz, aynı zamanda müşteri tercihlerini, pazar trendlerini ve toplumsal eğilimleri anlamak için analiz edilebilecek önemli bilgilere dönüşür.

Örneğin, Facebook, Instagram ve Twitter gibi sosyal medya platformları, her gün milyarlarca fotoğraf, video, metin ve etkileşimin paylaşıldığı devasa veri kaynaklarıdır. Bu tür veriler, işletmelerin pazarlama stratejilerini hedef kitlenin gerçek zamanlı davranışlarına göre şekillendirmelerine olanak tanır (Zeng, Proctor, & Li, 2013).

IoT cihazları, hem endüstriyel hem de bireysel düzeyde veri üretiminde büyük bir artışa yol açmaktadır. Akıllı evler, giyilebilir cihazlar, sağlık izleme cihazları, akıllı araçlar ve üretim makineleri gibi cihazlar, sürekli olarak veri üreterek işletmelerin büyük veri kümeleriyle başa çıkmasını zorlaştırmaktadır. Bu cihazlardan elde edilen verilerin her biri, genellikle yapılandırılmamış veri türlerinden oluşur ve analiz edilmeden önce işlenmesi gerekir. Cisco'nun 2020 yılına ilişkin raporuna göre, IoT cihazları aracılığıyla dünya genelinde üretilen veri hacmi, 2025 yılı itibarıyla yaklaşık 80 zettabayt (ZB) olacak ve bu, veri yönetiminde büyük bir zorluk yaratacaktır (Cisco, 2018).

Veri hacminin artması, veri depolama ve işleme altyapılarının da genişletilmesini gerektirir. Geleneksel veritabanları, bu kadar büyük veri hacmini verimli bir şekilde saklamak ve işlemek için yetersiz kalabilir. Bu noktada, bulut bilişim (cloud computing) ve dağıtık veri işleme sistemleri gibi yenilikçi teknolojiler devreye girer. Apache Hadoop gibi açık kaynaklı dağıtık işlem platformları, büyük veri setlerini paralel olarak işleyebilen ve verileri dağıtık bir şekilde saklayabilen sistemler sunar (Shvachko et al., 2010). Ayrıca, Amazon Web Services (AWS), Google Cloud ve Microsoft Azure gibi bulut servis sağlayıcıları, büyük veri kümesini barındırmak için esnek ve ölçeklenebilir çözümler sunmaktadır.

Büyük veri hacmi sadece depolama ile değil, aynı zamanda bu verilerin işlenmesiyle ilgili sorunları da beraberinde getirir. Veri akışı (streaming data), internet üzerindeki cihazlardan sürekli olarak veri toplama anlamına gelir ve bu tür verilerin gerçek zamanlı işlenmesi işletmeler için kritik hale gelir. Örneğin, Finansal hizmetler sektöründe yapılan ticaret işlemleri veya perakende sektöründe yapılan müşteri etkileşimleri, çok yüksek hızda gerçekleşen veri akışları yaratır. Bu tür verilerin hızlı bir şekilde işlenmesi, bir sonraki adımda alınacak kararları etkileyecektir (Berson, Smith, & Thearling, 2000).

Büyük veri hacminin yönetilmesinde önemli bir diğer zorluk da veri kalitesi meselesidir. Hacmi yüksek veri kümelerinde hatalı, eksik ya da yanlış etiketlenmiş veriler sıklıkla bulunur. Veri kalitesinin sağlanması, doğru analizler yapabilmek için çok önemlidir (Redman, 2013a). Bu nedenle, veriyi işleyen sistemlerin yüksek doğrulukla işlem yapabilmesi için büyük veri temizliği ve doğrulama süreçlerine yatırım yapılması gerekir.

Veri hacminin gelecekte daha da artacağı tahmin edilmektedir. IDC tarafından yapılan bir tahmine göre, 2025 yılı itibarıyla dünya çapında veri üretimi, her yıl %61 oranında artış gösterecek ve toplamda 175 zettabayt veri üretilecektir (IDC, 2020). Bu artış, işletmelerin yalnızca veriyi toplamak ve saklamakla kalmayıp, aynı zamanda bu verileri analiz etmek ve içgörüler üretmek için daha gelişmiş yapay zeka (AI) ve makine öğrenimi (ML) teknolojilerine gereksinim duyacağını göstermektedir. Ayrıca, 5G teknolojisi ile mobil internet hızlarının artması, daha fazla cihazın veri üretmesini ve bu verilerin daha hızlı bir şekilde işlenmesini sağlayacaktır.

Tabii, büyük verinin diğer özelliklerini detaylandırarak benzer şekilde genişletmek gerekmektedir. İzleyen paragraflarda Büyük Verinin Diğer Özellikleri üzerine detaylı bir açıklama yer alacaktır.

2.2. Çeşit (Variety)

Çeşit, büyük verinin en önemli özelliklerinden biridir. Bu, verinin farklı formatlarda, kaynaklardan ve türlerden gelmesi anlamına gelir. Verinin çeşitliliği, veri analizini ve işlenmesini karmaşık hale getiren unsurlardan birisidir. Geleneksel veritabanları, genellikle yapılandırılmış verilere dayanır. Ancak büyük veri, yapılandırılmış, yapılandırılmamış ve yarı yapılandırılmış veri türlerinin bir arada bulunduğu, çok daha çeşitli veri kümelerini içerir.

Yapılandırılmış veriler, geleneksel veritabanlarında kullanılan verilerdir. Bu veriler genellikle sayılar, tarih bilgileri veya metin gibi düzenli formatlarda bulunur. Örneğin, bir şirketin müşteri veritabanı, ürün satış bilgileri, fatura kayıtları ve envanter verileri yapılandırılmış veriler arasındadır. Bu tür veriler, relasyonel veritabanı yönetim sistemleri (RDBMS) ile kolayca işlenebilir. SQL dilinde yapılan sorgular ile bu veriler hızlıca analiz edilebilir (Codd, 1990).

Büyük verinin çoğu, yapılandırılmamış verilerden oluşur. Yapılandırılmamış veriler, verinin düzenli bir yapıya veya formatta olmaması durumudur. Bu tür veriler arasında sosyal medya gönderileri, videolar, sesli mesajlar, blog yazıları, e-posta içerikleri, web sayfaları, kullanıcı yorumları ve haber bültenleri yer alır. Veri madenciliği (data mining) ve doğal dil işleme (NLP) gibi yöntemler, yapılandırılmamış verilerin anlamlı hale getirilmesinde kullanılır. Bu verilerin analiz edilmesi ve işlenmesi daha karmaşık ve zaman alıcıdır (Manning et al., 2008).

Yarı yapılandırılmış veriler, belirli bir düzene sahip ancak tam olarak yapılandırılmış olmayan verilerdir. XML, JSON veya CSV gibi dosya formatları yarı yapılandırılmış verilere örnek olarak verilebilir. Bu veriler, bazı düzenlere ve kurallara sahiptir, ancak tam olarak ilişkisel veritabanlarında depolanabilir ve işlenebilir bir yapıda değildir. Yarı yapılandırılmış veriler, özellikle web verileri, cihaz verileri ve log dosyalarında yaygın olarak bulunur (Bizer et al., 2012).

Çeşitli veri türlerinin yönetilmesi, büyük veri analitiğinde önemli zorluklardan birini oluşturur. İşletmelerin farklı veri türlerini etkin bir şekilde entegre etmeleri ve anlamlı analizler çıkarmaları gerekir. Bu da, veri entegrasyonu, veri işleme ve analiz teknikleri konusunda gelişmiş teknolojilere yatırım yapmayı gerektirir. Veri çeşitliliği ile başa çıkabilmek için NoSQL veritabanları gibi modern çözümler, verilerin depolanması ve yönetilmesinde sıklıkla tercih edilmektedir (Stonebraker et al., 2010).

Veri çeşitliliği, IoT cihazları, sosyal medya, video içerikleri ve gelişen sensör teknolojileri ile daha da artmaktadır. Gelecekte, işletmelerin bu verileri entegre etme ve analiz etme yeteneği, rekabet avantajlarını belir-

leyecek önemli bir faktör olacaktır. Ayrıca, veri gölleri (data lakes) ve veri ambarları (data warehouses) gibi yeni veri depolama çözümleri, çok çeşitli veri türlerini tek bir yerde toplayarak daha etkili analizler yapılmasına olanak tanır (Gartner, 2019).

2.3. Hız (Velocity)

Hız, büyük verinin işlenme ve iletilme hızını ifade eder. Verinin hızlı bir şekilde toplanması, analiz edilmesi ve sonuçların çıkarılması, özellikle gerçek zamanlı veri akışlarıyla çalışan işletmeler için kritik öneme sahiptir. Gerçek zamanlı analiz (real-time analytics), akıllı şehirler, finansal hizmetler, sağlık sektörü ve perakende gibi birçok endüstride işletmelerin karar verme süreçlerini hızlandırır.

Büyük veri hızının en dikkat çekici özelliği, verilerin sürekli olarak üretilmesi ve anlık olarak işlenmesidir. Örneğin, finansal ticaret sistemleri, hisse senedi işlemleri gibi finansal hizmetlerde, piyasadaki her küçük değişiklik anında takip edilmekte ve bu veriler gerçek zamanlı olarak işlenmektedir. Hızlı veri işleme gereksinimi, çok kısa süreli fiyat değişimlerini analiz etmek ve bu değişimlere hızlı bir şekilde tepki vermek için kritik bir öneme sahiptir (Samarati et al., 2016).

Veri hızının gelecekte daha da artması beklenmektedir. 5G teknolojinin yaygınlaşması, verinin daha hızlı iletilmesini sağlarken, daha fazla cihazın birbirine bağlanmasına olanak tanıyacaktır. Bu, büyük verinin hızla üretilmesi ve işlenmesi gereksinimini daha da önemli hale getirecektir. Ayrıca, yapay zeka ve makine öğrenimi algoritmaları, büyük veri hızını analiz etmede daha etkili yöntemler geliştirecek ve işletmelere hız avantajı sağlayacaktır.

Yukarıda sıralanan özelliklere ek olarak bahsedilmesi gereken bir diğer kavram Doğruluktur. Doğruluk, büyük verinin güvenilirliğini ve doğruluğunu ifade eder. Büyük verinin kullanımı, doğruluk sorunlarıyla karşılaşılabilir, çünkü çok büyük veri kümelerinde hatalı, eksik veya yanıltıcı veriler bulunabilir. Veri doğruluğu, karar alma süreçlerinde yanlış sonuçlara yol açabilir, bu da işletmelerin yanlış stratejik kararlar almasına neden olabilir. Bu nedenle, doğruluğun sağlanması, büyük veri analitiğinin kritik bir parçasıdır. Veri kalitesinin düşük olması, işletmelerin etkili veri analizi yapmasını engeller. Veri doğruluğu, verilerin geçerliliği ve güvenilirliği ile doğrudan ilgilidir. Örneğin, müşteri bilgileri içeren verilerde eksik veya yanlış bilgiler, işletmenin pazarlama stratejilerini yanlış yönlendirebilir. Veri doğruluğu sağlanması için veri temizleme (data cleaning) ve doğrulama (validation) yöntemleri gereklidir (Redman, 2013b).

3. VERİ ANALİTİĞİ: TANIM VE TÜRLERİ

Veri analitiği, büyük verinin anlamlı ve değerli bilgilere dönüştürülmesi sürecini ifade eder. Bu süreç, genellikle verilerdeki desenleri, ilişkileri, trendleri ve anormallikleri belirlemeye yönelik çeşitli teknikler ve algoritmalar içerir. Veri analitiği, işletmelerin daha bilinçli kararlar almasını, operasyonel verimliliklerini artırmalarını, müşteri deneyimlerini iyileştirmelerini ve stratejik hedeflerine ulaşmalarını sağlar. Bu süreçte kullanılan araçlar ve teknikler, verinin türüne, amaca ve analiz edilecek probleme göre değişir. Veri analitiği, genellikle **tanımlayıcı**, **öngörücü**, **preskriptif** ve **algoritmik** olmak üzere dört ana türe ayrılır.

3.1. Tanımlayıcı Analitik (Descriptive Analytics)

Tanımlayıcı analitik, geçmiş verilere dayalı olarak ne olduğunu anlamaya yönelik analiz türüdür. Bu tür analizler, geçmişteki olayları veya süreçleri inceleyerek, bir durumu açıklamak veya mevcut eğilimleri anlamak amacıyla yapılır. Tanımlayıcı analizler, verileri toplar, organize eder ve özetler, böylece kullanıcılar ve işletmeler geçmiş performansı değerlendirebilir. Örnek olarak;

- **Satış raporları:** Bir işletme, geçmiş aylar veya yıllar bazında satış performansını inceleyebilir.
- **Müşteri memnuniyeti anketleri:** Müşteri geri bildirimlerini analiz ederek, müşteri memnuniyeti trendlerini gözlemlemek.

Tanımlayıcı analitik, genellikle istatistiksel analiz, veri madenciliği ve özetleyici istatistikler gibi yöntemler kullanarak uygulanır. Bu analiz türü, genellikle veri görselleştirme araçlarıyla (örneğin, grafikler ve tablolar) sunulur. İşletmeler, tanımlayıcı analitik ile geçmiş verilerini analiz ederek performans göstergelerini izler ve bu bilgilerle gelecekteki stratejiler için bir temel oluştururlar. Örneğin, bir perakende şirketi, müşteri segmentasyonunu ve alışveriş alışkanlıklarını inceleyerek, hedef kitlelerini daha iyi tanıyabilir. Bu tür analizler, ayrıca işletmelerin operasyonel verimliliğini değerlendirmelerine ve kaynakları daha etkili kullanmalarına yardımcı olabilir (Chaudhuri et al., 2011).

3.2. Öngörücü Analitik (Predictive Analytics)

Öngörücü analitik, gelecekte ne olacağına dair tahminlerde bulunmayı amaçlayan analiz türüdür. Bu tür analizler, geçmiş veriler ve makine öğrenimi algoritmaları kullanarak gelecekteki eğilimleri veya olayları tahmin eder. Öngörücü analitik, verilerdeki örüntüleri tanımak ve bu örüntü-

lere dayalı olarak geleceği modellemek için çeşitli istatistiksel ve makine öğrenimi tekniklerini kullanır. Örnek olarak;

- **Müşteri davranışlarını tahmin etmek:** Bir e-ticaret sitesi, müşterilerin gelecekteki alışveriş davranışlarını tahmin etmek için geçmiş satın alma verilerini analiz edebilir.
- **Risk analizi:** Bir bankacılık kuruluşu, kredi başvurusu yapan kişilerin kredi ödeme olasılıklarını tahmin etmek için öngörücü analitik kullanabilir.

Öngörücü analitik, genellikle zaman serisi analizi, regresyon analizi, karar ağaçları ve makine öğrenimi gibi tekniklerden yararlanır. Bu tür analizler, işletmelere gelecekteki eğilimleri ve riskleri anlamada yardımcı olur. Öngörücü analitik, işletmelerin stratejik kararlarını daha sağlam temellere dayandırmalarını sağlar. Örneğin, perakende şirketleri, stok yönetimini optimize etmek ve talep tahminleri yapmak için öngörücü analitiği kullanabilir. Ayrıca, sağlık sektöründe, hastaların gelecekteki sağlık risklerini tahmin etmek ve tedavi planları oluşturmak için öngörücü analizler yapılmaktadır (Shmueli et al., 2017).

3.3. Preskriptif Analitik (Prescriptive Analytics)

Preskriptif analitik, gelecekteki olaylara dair önerilerde bulunarak, en iyi aksiyonları belirlemeyi amaçlayan analiz türüdür. Bu tür analizler, öngörücü analizlerden elde edilen tahminler üzerine kurulur ve bu tahminlere dayalı olarak en iyi kararları almak için tavsiyelerde bulunur. Preskriptif analitik, genellikle optimizasyon teknikleri, simülasyonlar ve algoritmalar kullanarak karar süreçlerini destekler. Örnek olarak;

- **Üretim optimizasyonu:** Bir üretim şirketi, en verimli üretim planını belirlemek için preskriptif analitiği kullanabilir.
- **Pazarlama kampanyaları:** Bir işletme, en etkili pazarlama stratejisini belirlemek için müşteri verilerinden elde edilen tahminleri ve önerileri kullanabilir.

Preskriptif analitik, genellikle lineer programlama, karmaşık optimizasyon algoritmaları ve simülasyon teknikleri gibi yöntemler içerir. Bu analiz türü, işletmelerin veri odaklı kararlar alarak operasyonel verimliliği en üst düzeye çıkarmasına yardımcı olur (Davenport & Harris, 2007). Preskriptif analitik, işletmelere karmaşık kararlar alırken yardımcı olur ve optimum sonuçları elde etmelerini sağlar. Örneğin, bir lojistik şirketi, teslimat rotalarını optimize etmek için preskriptif analiz teknikleri kullanarak maliyetleri düşürebilir. Benzer şekilde, sağlık sektöründe, doktorlar hasta-

lara tedavi seçeneklerini sunmadan önce en iyi tedavi yöntemini belirlemek için preskriptif analitik kullanabilirler.

3.4. Algoritmik Analitik (Algorithmic Analytics)

Algoritmik analitik, verileri analiz etmek için özel algoritmaların kullanıldığı bir analiz türüdür. Algoritmik analiz, özellikle büyük veri kümelerinde, karmaşık ilişkileri keşfetmek ve desenleri tanımak için kullanılır. Makine öğrenimi algoritmaları ve yapay zeka teknikleri, algoritmik analizde yaygın olarak kullanılır. Algoritmik analizler, genellikle çok büyük ve karmaşık veri kümeleriyle çalışırken, verilerin anlamlı hale getirilmesi için otomatikleştirilmiş yöntemler kullanır. Örnek olarak;

- **Yüz tanıma:** Güvenlik sistemlerinde, görüntü verilerinin analiz edilmesi için algoritmik analitik kullanılır.
- **Doğal dil işleme (NLP):** Metin verilerinin analiz edilmesi ve anlamlı sonuçlar çıkarılması için algoritmalar kullanılır.

Algoritmik analizler, genellikle derin öğrenme, karar ağaçları, klasifikasyon ve regresyon algoritmaları gibi yapay zeka ve makine öğrenimi tekniklerine dayanır. Algoritmik analitik, büyük verinin işlenmesi ve anlamlı hale getirilmesi için gereklidir. Örneğin, finansal hizmetlerde, algoritmalar borsa hareketlerini tahmin etmek için kullanılabilir. Benzer şekilde, sosyal medya platformlarında, kullanıcıların ilgi alanlarını anlamak için algoritmik analiz teknikleri kullanılır (Domingos, 2015).

Özetle, veri analitiği, işletmelerin verilerden anlamlı bilgiler çıkararak stratejik ve operasyonel kararlar almasını sağlar. Tanımlayıcı, öngörücü, preskriptif ve algoritmik analitik gibi farklı analiz türleri, işletmelerin veri odaklı kararlar almasına ve rekabet avantajı elde etmesine olanak tanır. Her bir analiz türü, belirli bir amaca hizmet eder ve farklı veri kümeleri ile çalışır. Bu nedenle, işletmelerin ihtiyaçlarına ve veri türlerine uygun analiz türlerini seçmeleri önemlidir.

4. İŞLETMELER İÇİN BÜYÜK VERİNİN ÖNEMİ

Büyük veri, işletmelerin stratejik hedeflerine ulaşmalarını sağlamak için kritik bir kaynak haline gelmiştir. Her gün üretilen veri miktarının artması, işletmelerin daha bilinçli kararlar almasını, operasyonel verimliliklerini artırmasını ve rekabet avantajı elde etmelerini mümkün kılmaktadır. Bu bölümde, büyük verinin işletmeler için önemi, sağladığı fırsatlar ve karşılaşılan zorluklar ele alınacaktır. Büyük verinin iş dünyasındaki yeri ve etkisi, sektörden sektöre değişiklik göstermekle birlikte, genel olarak tüm işletmeler için stratejik bir öneme sahiptir.

4.1. Rekabet Avantajı Sağlamak

Büyük veri, işletmelere rekabet avantajı elde etmek için kullanılabilircek güçlü bir araçtır. İşletmeler, verileri analiz ederek müşteri ihtiyaçlarını daha doğru bir şekilde anlayabilir, piyasa eğilimlerini tahmin edebilir ve bu sayede rakiplerinden bir adım önde olabilirler. Veri odaklı kararlar almak, işletmelere daha hızlı ve etkili tepki verme yeteneği kazandırır. Örneğin, perakende sektöründe, bir mağaza, müşterilerinin alışveriş alışkanlıklarını analiz ederek, doğru ürünleri doğru zamanda sunabilir. Bu tür bir veri analitiği, satışları artırabilir ve müşteri memnuniyetini iyileştirebilir (McAfee & Brynjolfsson, 2012).

Ayrıca, büyük veri analitiği, işletmelerin kaynakları daha verimli kullanmalarını sağlar. Bir üretim tesisinde, makinelerin verimliliğini izlemek ve bakım ihtiyaçlarını tahmin etmek, üretim süreçlerinin optimize edilmesine yardımcı olabilir. Bu da maliyetleri düşürür ve verimliliği artırır. İşletmeler, büyük veri sayesinde operasyonel süreçlerini daha esnek ve çevik hale getirebilir, bu da onlara hızlı değişen pazarlarda rekabet üstünlüğü sağlar (Davenport, 2014).

4.2. Müşteri Deneyimini İyileştirmek

Büyük veri, işletmelerin müşteri deneyimlerini özelleştirmelerini ve iyileştirmelerini sağlar. Müşteri verisi toplamak ve analiz etmek, bir işletmenin müşterilerinin davranışlarını, tercihlerine göre kararlarını daha doğru tahmin etmesine olanak tanır. Örneğin, e-ticaret platformları, kullanıcıların önceki alışverişlerine dayalı öneriler sunarak daha kişiselleştirilmiş bir deneyim yaratabilir. Bu tür kişiselleştirilmiş hizmetler, müşteri sadakatini artırabilir ve tekrar eden satışları teşvik edebilir. Büyük veri, aynı zamanda müşteri memnuniyetini izleme ve geri bildirimleri analiz etme konusunda da yardımcı olur. Müşterilerin sosyal medya üzerinden yaptıkları yorumlar, anketler ve şikayetler, işletmelerin hangi alanlarda iyileştirme yapmaları gerektiğini anlamalarına yardımcı olabilir. Müşteri memnuniyeti anketlerinden elde edilen verilerle, işletmeler, hizmet kalitesini iyileştirmek için doğru aksiyonları alabilir (Hofacker et al., 2016).

4.3. Karar Destek ve İyileştirme

Büyük veri analitiği, işletmelerin stratejik kararlarını daha bilinçli ve objektif bir şekilde almalarını sağlar. Veriler, işletmelerin performansını, pazar eğilimlerini, finansal durumu ve müşteri davranışlarını analiz ederek, yöneticilerin karar süreçlerine rehberlik eder. İşletmeler, büyük veri analitiğini kullanarak daha doğru tahminlerde bulunabilir ve bu tahminler doğrultusunda stratejik yönlerini belirleyebilirler.

Örneğin, bir şirket, büyük veri analitiği ile hangi ürünlerin daha fazla talep gördüğünü, hangi bölgelerde daha fazla pazar payı kazanabileceğini belirleyebilir. Bu veriler, işletmeye pazar stratejilerini, üretim miktarlarını ve reklam bütçelerini optimize etme fırsatı sunar. Ayrıca, büyük veri, maliyetleri izleme, bütçe tahminleri yapma ve yatırım geri dönüşlerini değerlendirme konusunda da kritik bir rol oynar (Mayer-Schönberger & Cukier, 2013).

4.4. İnovasyon ve Ürün Geliştirme

Büyük veri, işletmelerin yeni ürünler geliştirmelerini ve mevcut ürünlerini iyileştirmelerini sağlar. Müşterilerden, ürün kullanım verilerinden ve pazar araştırmalarından elde edilen veriler, yeni ürün fikirlerinin geliştirilmesinde temel bir kaynak oluşturur. Büyük veri analitiği ile işletmeler, müşteri taleplerini ve pazar boşluklarını daha iyi anlayarak yenilikçi ürünler yaratabilirler. Örneğin, bir otomobil üreticisi, araç kullanıcılarının sürüş alışkanlıkları, yol koşulları ve güvenlik ihtiyaçları gibi verileri analiz ederek, daha güvenli ve verimli araçlar tasarlayabilir. Ayrıca, büyük veri analitiği, ürünlerin performansını izlemek ve müşteri geri bildirimlerine göre ürün iyileştirmeleri yapmak için de kullanılabilir (Westerman et al., 2014).

4.5. Risk Yönetimi ve Güvenlik

Büyük veri analitiği, işletmelerin riskleri daha etkin bir şekilde yönetmelerini sağlar. Veriler, potansiyel risklerin erken tespiti için kullanılabilir. Örneğin, finansal hizmetlerde, büyük veri analitiği kullanılarak dolandırıcılık faaliyetleri, kredi riski ve yatırım riskleri analiz edilebilir. Ayrıca, büyük veri, operasyonel riskleri ve güvenlik açıklarını belirlemek için de kritik bir araçtır. Özellikle siber güvenlik alanında, büyük veri analitiği, güvenlik tehditlerini tespit etmek ve önlemek için kullanılmaktadır. Anormalliklerin izlenmesi, tehditlerin erken uyarı sistemleriyle tespit edilmesi, güvenlik önlemlerinin alınmasını kolaylaştırır. Bu, işletmelere daha güvenli bir çalışma ortamı sunar ve veri kaybı veya diğer zararlardan korunmalarına yardımcı olur (Barton & Court, 2012).

Yukarıda yazılanlardan anlaşılmaktadır ki; Büyük veri, işletmelerin stratejik kararlar alırken, operasyonel süreçlerini iyileştirirken, müşteri deneyimlerini özelleştirirken ve yenilikçi ürünler geliştirme konusunda kritik bir rol oynar. İşletmeler, büyük veri sayesinde yalnızca geçmiş verileri değil, aynı zamanda gelecekteki eğilimleri ve riskleri de tahmin edebilir. Ancak, büyük veri kullanımı, aynı zamanda çeşitli zorlukları da beraberinde getirmektedir. Veri güvenliği, veri kalitesi ve doğru analiz araçlarının seçilmesi gibi konular, büyük verinin etkin kullanımı

için önemli engeller oluşturabilir. Bu nedenle, işletmelerin büyük veri stratejilerini dikkatlice planlaması ve yönetmesi gerekmektedir.

5. BÜYÜK VERİNİN ZORLUKLARI

Büyük veri, işletmeler için önemli fırsatlar sunsa da, bunun yanında bir dizi zorluk da beraberinde getirir. Bu zorluklar, hem teknik hem de operasyonel düzeyde olabilir. Verinin doğru bir şekilde toplanması, saklanması, işlenmesi ve analiz edilmesi, büyük veriden tam anlamıyla yararlanabilmek için önemli bir dizi engel teşkil eder. Bu bölümde, büyük verinin işletmelere sunduğu fırsatların yanı sıra karşılaşılan başlıca zorluklar detaylı olarak ele alınacaktır.

5.1. Veri Kalitesi

Büyük veri projelerinde karşılaşılan en büyük zorluklardan biri veri kalitesidir. Verinin kalitesi, elde edilen verinin doğruluğu, tutarlılığı, güvenilirliği ve geçerliliği ile doğrudan ilişkilidir. Büyük veri, genellikle çok çeşitli kaynaklardan elde edilir; sosyal medya, sensörler, işlem kayıtları gibi veri kaynakları, bazen hatalı, eksik veya tutarsız veriler içerebilir. Bu durum, analizlerin doğru sonuçlar üretmesini zorlaştırabilir ve karar alıcıların yanıltılmasına yol açabilir. Veri kalitesinin düşük olması, özellikle tahminsel analizler ve karar destek sistemleri gibi alanlarda ciddi sorunlar yaratabilir. Örneğin, yanlış veya eksik verilerle yapılan bir müşteri segmentasyonu çalışması, hatalı stratejilere ve müşteri memnuniyetsizliğine neden olabilir. Bu nedenle, büyük veri projelerinin başarılı olabilmesi için veri kalitesinin sürekli olarak izlenmesi ve iyileştirilmesi gerekir. Verilerin doğru şekilde temizlenmesi, düzenlenmesi ve doğruluğunun sağlanması, veri analitiği süreçlerinin başarısını doğrudan etkileyen bir faktördür (Redman, 2013a).

5.2. Veri Güvenliği ve Gizliliği

Büyük verinin bir başka önemli zorluğu, veri güvenliği ve gizliliğidir. İşletmeler, büyük veri toplarken, müşteri bilgileri, finansal veriler ve kişisel veriler gibi hassas verilerle ilgilenmektedir. Bu verilerin doğru bir şekilde korunması, hem yasal düzenlemelere uygunluk sağlamak hem de müşterilerin güvenini korumak açısından kritik öneme sahiptir. Veri ihlalleri, yalnızca işletmelerin itibarını zedelemekle kalmaz, aynı zamanda büyük para cezalarına ve yasal sorunlara da yol açabilir. Özellikle GDPR (General Data Protection Regulation) gibi düzenlemeler, kişisel verilerin korunmasına dair katı kurallar getirmiştir ve büyük veri yönetimini daha karmaşık hale getirmiştir. Bu tür yasalar, işletmelerin veri güvenliğini sağ-

lamak için daha katı güvenlik önlemleri almalarını gerektirmektedir. Ayrıca, verilerin güvenli bir şekilde saklanması ve iletilmesi, işletmelerin daha iyi bir siber güvenlik stratejisi geliştirmelerini gerektirir (Mayer-Schönberger & Cukier, 2013).

5.3. Veri Yönetimi ve Entegrasyonu

Büyük veri projeleri genellikle farklı veri kaynaklarından gelen verilerin toplanmasını gerektirir. Bu veri kaynakları, işletmenin iç sistemlerinden, dış kaynaklardan veya sensörlerden gelen veriler olabilir. Ancak, farklı veri türlerinin birleştirilmesi ve entegre edilmesi, büyük bir zorluk oluşturabilir.

Veri entegrasyonu, özellikle verilerin farklı formatlarda (yapısal, yarı yapısal ve yapılandırılmamış) olması durumunda karmaşık hale gelir. Verilerin doğru bir şekilde entegre edilmesi, analiz süreçlerinde kullanılacak bütüncül ve tutarlı bir veri seti oluşturulmasını sağlar. Ayrıca, veri yönetimi, büyük veri projelerinde kaynakların etkin bir şekilde kullanılması için kritik bir faktördür. Doğru bir veri yönetimi stratejisi oluşturmak, veri kalitesini iyileştirir ve analizlerin doğruluğunu artırır (Laudon & Laudon, 2016).

5.4. Analitik ve Yetenek Gelişimi

Büyük verinin analizi, ileri düzeyde analitik beceriler ve güçlü teknoloji altyapısı gerektirir. Veri bilimcileri, veri analistleri ve diğer teknik uzmanlar, büyük veri kümelerini işleyebilmek ve anlamlı çıkarımlar yapabilmek için derinlemesine bilgiye sahip olmalıdır. Ancak, bu tür yetenekler çok talep görmekte ve sektördeki nitelikli iş gücü sıkıntısı, büyük veri projelerinin uygulanabilirliğini zorlaştırmaktadır. Ayrıca, büyük veri analizinin karmaşıklığı, analitik süreçlerde hataların yapılmasına yol açabilir. Yanlış analizler veya eksik veriler, yanlış sonuçlar doğurabilir ve stratejik kararları olumsuz yönde etkileyebilir. Bu nedenle, işletmelerin büyük veri analiz kapasitesini güçlendirecek eğitim programları geliştirmeleri ve doğru araçları kullanmaları gerekmektedir. Bu, analitik çözümlerin doğruluğunu ve güvenilirliğini artırır (Provost & Fawcett, 2013).

5.5. Yüksek Maliyetler

Büyük veri projelerinin uygulanması, özellikle başlangıç aşamasında yüksek maliyetlere yol açabilir. Gerekli teknoloji altyapısının kurulması, veri toplama ve depolama süreçlerinin yönetilmesi, yazılım ve donanım yatırımları gibi maliyetler, büyük veri projelerini küçük ve orta ölçekli işletmeler için zorlayıcı hale getirebilir. Ayrıca, veri analitiği için kullanılan

araçların, yazılımların ve sistemlerin geliştirilmesi de işletmelere ek maliyet yükü getirebilir. Büyük veri altyapısının kurulması ve işletilmesi, aynı zamanda sürekli bakım ve güncelleme gerektirir. Veri yönetimi ve güvenlik stratejilerinin sürdürülebilir olması için yapılan yatırımlar da önemli maliyet unsurlarındandır. Bu nedenle, büyük veri projelerinin uygulanabilirliğini değerlendirirken, maliyet-fayda analizi yapmak, işletmelerin stratejik kararlarını yönlendirebilir (Hadoop, 2015).

5.6. Veri Hacmi ve Depolama Zorlukları

Büyük veri, büyük hacimli veri kümelerinin işlenmesini ve depolanmasını gerektirir. Veri hacmi arttıkça, bu verilerin saklanması, yedeklenmesi ve erişilmesi daha karmaşık hale gelir. Verilerin doğru bir şekilde depolanması, verinin hızlı bir şekilde erişilebilmesini ve analiz edilmesini sağlar. Ancak, büyük veri depolama altyapısının kurulması, yüksek maliyetlere ve teknik zorluklara yol açabilir. Veri depolama zorlukları, özellikle verinin büyüklüğü arttıkça daha belirgin hale gelir. Verinin hızla artan boyutları, işletmeleri depolama kapasitesi konusunda sınırlarla karşı karşıya bırakabilir. Bu nedenle, işletmeler, büyük veri için ölçeklenebilir depolama çözümleri geliştirmeli ve bu veriyi işleyebilecek güçlü altyapılar kurmalıdır (Marz & Warren, 2015).

Görülmektedir ki; Büyük veri, işletmelere birçok fırsat sunarken, aynı zamanda çeşitli zorlukları da beraberinde getirmektedir. Veri kalitesi, güvenlik, entegrasyon, analitik beceri eksiklikleri, yüksek maliyetler ve depolama sorunları, işletmelerin büyük veriyi etkili bir şekilde kullanmalarını engelleyen başlıca engellerdir. Bu zorluklarla başa çıkabilmek için işletmelerin güçlü bir strateji geliştirmeleri, doğru teknolojilere yatırım yapmaları ve veri yönetimi süreçlerini sürekli olarak iyileştirmeleri gerekmektedir.

6. SONUÇ

Büyük veri, günümüz iş dünyasında stratejik bir kaynak haline gelmiştir. Teknolojinin ve dijitalleşmenin hızla ilerlemesiyle birlikte, işletmeler, büyük veri analitiği sayesinde daha bilinçli kararlar alabilir, operasyonel verimliliklerini artırabilir ve müşteri deneyimlerini özelleştirebilir. Ancak, büyük verinin sunduğu fırsatlar, aynı zamanda bazı önemli zorlukları da beraberinde getirmektedir. Veri güvenliği, kalitesi, yönetimi, analitik beceri eksiklikleri ve yüksek maliyetler, büyük veri projelerinin etkin bir şekilde hayata geçirilmesinde karşılaşılan başlıca engellerdir.

Büyük verinin etkin bir şekilde kullanılması, yalnızca doğru teknolojilere yatırım yapmayı değil, aynı zamanda organizasyonel bir dönüşümü

gerektirir. Veri kültürünün benimsenmesi, şirket içinde veri analitiği konusunda uzmanlaşmış ekiplerin oluşturulması ve veri tabanlı karar verme süreçlerinin entegrasyonu, başarılı bir büyük veri stratejisi için kritik öneme sahiptir. İşletmeler, yalnızca veri toplamakla kalmayıp, aynı zamanda bu verileri anlamlı bir şekilde analiz ederek, karar alma süreçlerine entegre etmelidirler.

Büyük verinin geleceği, teknolojik gelişmelerle paralel olarak şekillenecektir. Yapay zeka, makine öğrenimi, nesnelerin interneti (IoT), bulut bilişim ve 5G teknolojilerinin hızla gelişmesi, büyük veri analizlerinin daha hızlı, daha doğru ve daha verimli yapılabilmesini mümkün kılacaktır. Bu gelişmeler, veri toplama ve işleme süreçlerini daha verimli hale getirerek, işletmelere daha kısa sürelerde değerli bilgiler sunacaktır.

Özellikle yapay zeka ve makine öğrenimi, büyük veri analitiği üzerinde büyük bir etkiye sahip olacak ve işletmelerin veriyi daha ileri düzeyde analiz etmelerini sağlayacaktır. Makine öğrenimi algoritmaları, büyük veri setlerinden otomatik olarak anlamlı desenler çıkarabilecek ve gelecekteki eğilimleri tahmin edebilecektir. Bu da işletmelerin sadece geçmiş verilere dayalı kararlar almakla kalmayıp, aynı zamanda gelecekteki pazar eğilimlerini, müşteri davranışlarını ve potansiyel fırsatları da tahmin etmelerine imkan verecektir (Brynjolfsson & McAfee, 2014).

Büyük veri projelerinin geleceği, bu verilerin etik bir şekilde kullanılmasıyla doğrudan ilişkilidir. İşletmelerin, veriyi toplayan ve işleyen sistemlerin etik standartlara uygun olmasını sağlamaları gerekmektedir. Ayrıca, veri analitiği ve yapay zeka gibi teknolojilerin kullanımında, algoritmaların önyargısız, adil ve şeffaf olmasına özen gösterilmelidir. Bu, hem yasal uyumluluğu sağlamak hem de müşteri güvenini korumak açısından kritik bir faktördür (O'Neil, 2016).

Küçük ve orta ölçekli işletmeler için büyük verinin geleceği, daha erişilebilir ve uygun maliyetli çözümler geliştirilmesiyle şekillenecektir. Örneğin, bulut tabanlı veri analitiği platformları, küçük işletmelere büyük veri analizlerini kolayca yapma imkanı tanıyacaktır. Bu işletmeler, pazarlama stratejilerinden müşteri ilişkilerine kadar geniş bir yelpazede büyük veri kullanarak, büyüme fırsatlarını keşfetme şansı elde edeceklerdir (Chong et al., 2017).

Büyük veri, işletmeler için yalnızca bir kaynak değil, aynı zamanda rekabet avantajı sağlayan güçlü bir stratejik araçtır. Veriyi anlamlı bir şekilde kullanabilen işletmeler, pazarlama, müşteri ilişkileri, tedarik zinciri yönetimi, ürün geliştirme ve daha birçok alanda üstünlük sağlayabilir. Bununla birlikte, büyük verinin yönetimi ve analizi konusunda karşılaşılan zorluklar, işletmeleri bu fırsatları en verimli şekilde kullanmaktan alıkoymaktadır.

Büyük verinin etkin kullanımı, işletmelerin daha esnek, yenilikçi ve müşteri odaklı olmasını sağlayacak ve onları gelecekteki dijital dönüşüm süreçlerine hazırlayacaktır. Ancak, bu dönüşüm süreci yalnızca teknolojik altyapının güçlendirilmesiyle değil, aynı zamanda organizasyonel değişimle mümkün olacaktır. İşletmeler, büyük veri kullanımında karşılaştıkları zorlukları aşarak, rekabetçi pazarlarda güçlü bir konum elde edebilirler.

KAYNAKÇA

- BARTON, D., & COURT, D. (2012). **Making Advanced Analytics Work for You**. Harvard Business Review, 90(10), 78-83.
- BERSON, A., SMİTH, S. J., & THEARLING, K. (2000). **Building Data Mining Applications for CRM**. McGraw-Hill.
- BİZER, C., BONCZ, P., & LAWLEY, M. (2012). **The Meaning of Big Data: Data Governance and Data Modeling**. Springer.
- BRYNJOLFSSON, E., & MCAFEE, A. (2014). **The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies**. W.W. Norton & Company.
- CHAUDHURİ, S., DAYAL, U., & NARASAYYA, V. (2011). **An Overview of Business Intelligence Technology**. Communications of the ACM, 54(8), 88-98.
- CHEN, H., CHİANG, R. H. L., & STOREY, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.
- CHONG, A. Y. L., ET AL. (2017). **The Impact of Big Data Analytics on Business Intelligence: A Comprehensive Review**. International Journal of Information Management, 37(6), 768-776.
- CİSCO. (2018). **Cisco Annual Internet Report (2018–2023)**. Retrieved from: <https://www.cisco.com>
- CODD, E. F. (1990). **The Relational Model for Database Management: Version 2**. Addison-Wesley.
- DAVENPORT, T. H. (2014). **Analytics at Work: Smarter Decisions, Better Results**. Harvard Business Review Press.
- DAVENPORT, T. H., & HARRİS, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business School Press.
- DOMİNGOS, P. (2015). **The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World**. Basic Books.
- GARTNER. (2012). *The 3 V's of Big Data*. Retrieved from <https://www.gartner.com/newsroom>
- GARTNER. (2019). **Magic Quadrant for Data Management Solutions for Analytics**. Gartner, Inc.
- HADOOP, A. (2015). **The Economics of Big Data**. International Journal of Computer Science & Technology, 3(1), 10-16.
- HOFACKER, C. F., MALTHOUSE, E. C., & SULTAN, F. (2016). **Big Data and the Customer Experience**. Journal of the Academy of Marketing Science, 44(2), 230-248.

- IDC. (2020). **Data Age 2025: The Digitization of the World**. International Data Corporation. Retrieved from: <https://www.idc.com>
- LAUDON, K. C., & LAUDON, J. P. (2016). **Management Information Systems: Managing the Digital Firm**. Pearson Education.
- MANNING, C. D., RAGHAVAN, P., & SCHÜTZE, H. (2008). **Introduction to Information Retrieval**. Cambridge University Press.
- MARZ, N., & WARREN, J. (2015). **Big Data: Principles and Paradigms**. Manning Publications.
- MAYER-SCHÖNBERGER, V., & CUKIER, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.
- MCAFEE, A., & BRYNJOLFSSON, E. (2012). **Big Data: The Management Revolution**. Harvard Business Review, 90(10), 60-68.
- MCKINSEY & COMPANY. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*.
- O'NEIL, C. (2016). **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. Crown Publishing Group.
- PROVOST, F., & FAWCETT, T. (2013). **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. O'Reilly Media.
- REDMAN, T. C. (2013). **Data Driven: Profiting from Your Most Important Business Asset**. Harvard Business Review Press.
- REDMAN, T. C. (2013). **Data Quality: The Accuracy Dimension**. Morgan Kaufmann.
- SAMARATI, P., BALDINI, G., & BLAZZINI, S. (2016). **Big Data and Real-Time Analytics: Techniques and Tools**. Springer.
- SHMUELI, G., PATEL, N. R., & BRUCE, P. C. (2017). **Data Mining for Business Analytics: Concepts, Techniques, and Applications in R**. Wiley.
- SHVACHKO, K., KUANG, H., RADIA, S., & CHANSLER, R. (2010). **The Hadoop Distributed File System**. Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies.
- STONEBRAKER, M., KEMNITZ, G., & O'NEIL, P. (2010). **The Design of the NoSQL Database Systems**. ACM Computing Surveys, 42(1), 1-43.
- WESTERMAN, G., BONNET, D., FERRARIS, P., & HEMERLING, J. (2014). **The Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry**. MIT Sloan Management Review and Capgemini Consulting.
- ZENG, D., PROCTOR, R., & LI, Y. (2013). **Social Media and Big Data Analytics for Business**. Wiley & Sons.

YÖNETİM BİLİŞİM SİSTEMLERİNDE İLERİ ARAŞTIRMALAR

yaz
yayınları

YAZ Yayınları
M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar / AFYONKARAHİSAR
Tel : (0 531) 880 92 99
yazyayinlari@gmail.com • www.yazyayinlari.com