



YÜKSEKÖĞRETİM KURULU BÜYÜK VERİ PROJESİ

Ankara - Mayıs 2022

YÜKSEKÖĞRETİM KURULU BÜYÜK VERİ PROJESİ

ANKARA ÜNİVERSİTESİ BASİMEVİ
İncitaşı Sokak No:10
06510 Beşevler / ANKARA
Tel: 0 (312) 213 66 55
Basım Tarihi:17.05.2022

İÇİNDEKİLER

I. Sunuş.....	5
II. Büyük Veri Projesi Çalışma Grubu	7
III. Büyük Veri Projesinde Pilot Üniversiteler.....	9
IV. Büyük Veri Çalışma Grubu İstanbul Toplantısı Görüşler ve Öneriler.....	11
V. Büyük Veri Çalışma Grubu Makaleleri	23
• Yükseköğretim Ekosisteminde Dijitalleşme ve Büyük Veri	25
• Yüksek Öğretimde Veri Yönetişimi	29
• Bilgi Çağında Büyük Verinin Önemi	33
• Eğitimde Büyük Veri Kullanımı	35
• Büyük Veri Öğretiminde İlk Adım: Gereksinimler ve Mimari Katmanlar	39
• Verimli Öğrenme İçin Büyük Veri Kullanan Arama ve Öneri Sistemleri	43
• Büyük Veri ve Veri Biliminin Yükseköğretimdeki Olası Kullanım Alanları.....	45

SUNUŞ



Değerli Arkadaşlarım,

Son 10 yılda, Bilgi Ekonomisi ve yeni teknolojilerin etkisi ile yükseköğretimde hızlı ve hacimli birçok yenilik gündeme gelmektedir. Küresel öğrenme, dünyamızın birbirine bağlılığını ve ülkelerin, toplumun ortak iyiliği için çok daha kapsayıcı bir akademik düşünceye sahip olmalarını gerektirmektedir. Bu bağlamdan olmak üzere, Büyük Veri çalışmaları, kurumlar için yeni teknolojilerin sağladığı en önemli kazanımlardan biridir.

İnternetin gelişmesi, sosyal medya kullanımının artması, sensör verileri, sunuculardan alınan log kayıtları, Büyük Veride ciddi artışlar sağlamaktadır. Uluslararası Veri Kurumu (IDC- International Data Corporation), dünyadaki verilerin her iki yılda bir ikiye katlandığını iddia etmektedir.

2001'de Büyük Veri kavramının boyut ve özelliklerinin 3V'si (volume, variety, velocity-hacim, çeşitlilik ve hız) Doug Laney tarafından tanımlandı. 2005 yılında Bilgisayar bilimcileri Doug Cutting ve Mike Cafarella ve ekip arkadaşları, büyük veri kümelerini depolamak ve işlemek için kullanılan açık kaynaklı yazılım araçlarını geliştirdiler. Bu çalışmalar, Big Data terminolojisi ile Roger Margoulas tarafından 2005 yılında literatüre kazandırıldı.

Geleneksel veri yönetimi tekniklerinin karmaşıklığı ve büyüklüğü nedeni ile verinin yönetiminde ve işlenmesindeki zorlukları tanımlamak ve bu zorlukları aşabilmek amacıyla ortaya konulmuş bir kavramdır. Başlangıçta bu alandaki çalışmalarda, Büyük Veri, yüksek hacimli, yüksek hızlı ve çok çeşitli olan verileri tanımlamak için kullanılan bir terim olarak ifade edilmektedir. 1990'lı yıllarda en büyük veri boyutu Terabyte ile ifade edilirken, günümüzde Zetabyte ve Zetabyte'ın bin katı olan Yotabyte ile ifade edilmektedir.

Yükseköğretim Kurulu, Türk Yükseköğretiminde Büyük Verinin inovatif anlamda eğitimde, veriyi toplama ve analiz etmede ve Büyük Veri ile ilgili en önemli bileşen olan Değer'in (Value) elde edilen, işlenen ve analiz edilen verilerin kurum için ciddi önem taşıyacağı bilinci doğrultusunda 2022'de, , pilot üniversite ile bir proje kapsamında konuyu çalışma kararı almıştır. Farklı kaynaklardan elde edilen büyük miktarda verinin keşfedilmesi ve yaygın kullanımı ulusal ve uluslararası düzeyde, Türk Yükseköğretimine önemli ivmeler kazandıracaktır.

Yükseköğretimde eğitimde ve araştırmada, Büyük Veri, eşsiz fırsatlar vadetmektedir. Ancak, Büyük Veri ile çalışan ve konuyu derinlemesine bilen profesyonel araştırmacıların sayısı gereğinden çok azdır. Büyük Veri ile çalışmak, çok iyi yetişmiş Veri Uzmanları ve Akademisyenleri ile mümkündür. Bu alanda çalışanların, istatistik, metodoloji ve görselleştirme bilgisine sahip ve ayrıca ortaya çıkabilecek ciddi etik çelişkiler ile ilgili donanımlı olmaları gerekmektedir.

Büyük Veri ile ilgili 2003-2022 yılları arasında Türkiye'de üniversitelerimizde, 57 doktora ve 171 yüksek lisans tezi olmak üzere 228 tez çalışması üretilmiştir (YÖK Tez Merkezi).

Büyük Veri Projemiz ve yükseköğretimde Büyük Veri Madenciliği çalışmaları, kanıta dayalı araştırma, öğrenme ve öğretme teknikleri konuları başta olmak üzere özellikle üniversitelerin araştırmacılarına ve araştırma yapısına ciddi anlamda bir gelişme vadetmektedir.

Özellikle son yıllarda Büyük Verinin, yükseköğretimin gelişmesine ve dönüşümüne olan katkısı gittikçe büyümektedir. Bu proje ile üniversitelerimizde Veri Bilimi çalışmalarına da daha fazla önem vermeyi hedefliyoruz.

Projeye başarılar diliyorum.

Erol ÖZVAR
YÖK Başkanı
Mayıs, 2022

BÜYÜK VERİ PROJESİ ÇALIŞMA GRUBU

Dr. Ali Taha Koç- Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanı

Doç. Dr. İhsan Tolga Medeni- Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Uzmanı

Prof. Dr. Erkay Savaş- Sabancı Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi Dekanı

Prof. Dr. Şule Gündüz Öğüdücü – İTÜ

Doç. Dr. Ahmet Cüneyd Tantuğ- İTÜ

Prof. Dr. Hasan Karal- Trabzon Üniversitesi

Doç. Dr. Ali Kürşat ERÜMIT – Trabzon Üniversitesi

Doç. Dr. Adnan Özsoy- Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü

Doç. Dr. İsmail Sengör Altıngövde- ODTÜ Bilgisayar Mühendisliği

Dr. Öğretim Üyesi Hamdi Dibeklioğlu- Bilkent Üniversitesi

BÜYÜK VERİ PROJESİNDE PİLOT ÜNİVERSİTELER

İstanbul Teknik Üniversitesi
Orta Doğu Teknik Üniversitesi
Atatürk Üniversitesi
İzmir Kâtip Çelebi Üniversitesi
Sabancı Üniversitesi
Fırat Üniversitesi
Sakarya Üniversitesi
Trabzon Üniversitesi

**BÜYÜK VERİ PROJESİ ÇALIŞMA GRUBU
İSTANBUL TOPLANTISI
GÖRÜŞLER ve ÖNERİLER
7 Nisan, 2022**

YÜKSEKÖĞRETİMDE BÜYÜK VERİ: DÜNYADAN UYGULAMA ÖRNEKLERİ VE TÜRKİYE'DE POTANSİYEL UYGULAMA ALANLARI

Orta Doğu Teknik Üniversitesi Görüş Raporu

Önceki akademik çalışmalar ışığında yükseköğretim kurumlarında toplanılan veri başlıca şu kategoriler altında değerlendirilebilir: Öğrenci verisi, eğitim ve öğrenme verisi, yönetsel veri, bölüm/müfredat verisi ve araştırma verisi. En genel ayrımla bu (büyük) veriler üzerinde iki çeşit veri analitiği (ve kestirimci modelleme) yapılması mümkündür: akademik/kurumsal analitik ve öğrenme analitiği [1, 2]. İlk gruptaki analitik çalışmaları kurum, bölge ve ülke seviyesinde uygulamalara odaklanırken, ikinci gruptaki öğrenme analitiği çalışmaları öncelikle öğrenci seviyesinde analiz ve uygulamalara odaklanmaktadır. Hanover Research şirketinin 2016 tarihli bir raporunda, bu iki çeşit veri analitiğinin çeşitli seviyelerde uygulandığı örnek vakaların bir kısmı aşağıdaki tablodaki gibi özetlenmiştir [1].

Analitik kategorisi	Analiz seviyesi	Analiz türü
Öğrenme Analitiği	Ders seviyesi	Kişiselleştirilmiş müfredat
		Öğrenci başarımlarını değerlendirme
	Öğrenci seviyesi	Başarımların kestirimi ve erken uyarı
		Otomatik danışmanlık
Bölüm seviyesi	Kestirimci modelleme ve erken uyarı	
Akademik/kurumsal analitik	Akademisyen seviyesi	Öğretme etkinliği
		Finansal katkı
	Öğrenci seviyesi	Kabul profili ve kestirimci analiz
		Okul sonrası istihdam analizi
	Alan veya ders önerisi	
	Kurumsal seviye	Kabul analizi
		Kurumsal başarımlar
		Okula devam

Tabloda sunulan analiz/uygulama yaklaşımlarını somutlaştırmak adına, yurtdışındaki çeşitli üniversitelerden seçilen ve öğrencileri hedefleyen büyük veri ve veri analitiği uygulama örnekleri ise şu şekilde özetlenebilir (2017 öncesi geliştirilen bazı uygulamalar halihazırda kullanımdan kaldırılmış olabilir):

- Course Signals (Purdue Üniv., ABD, 2012) [3]: Bu uygulamada öğrencinin genel başarı bilgileri ile belirli bir dersteki faaliyetleri hakkında öğrenme yönetim sisteminden (mesela, Moodle) alınan bilgileri kestirimci bir modele verilerek o dersten geçme/kalma olasılığı belirlenmekte ve dönem boyunca öğrenciye uyarı mesajları gönderilebilmektedir. Bu uygulamanın öğrencilerin ders başarısının artmasında etkili olduğu, dersten kalma riskini daha erken fark eden öğrencilerin yardım arama davranışının arttığı görülmüştür. Ayrıca özellikle ön-şart durumundaki derslerden kalma olasılığının azalmasıyla zamanında mezun olma oranında da artış gözlemlenmiştir. Ancak bu sonuçların geçerliliği üzerinde bazı tartışmalar da bulunmaktadır.

Burada anlatılan Course Signals gibi uygulamaların yaptığı bazı kestirimsel analizler, günümüzde Moodle gibi açık kaynaklı öğrenme yönetim sistemleri içerisinde de yapılabilir (mesela, Moodle yardım sayfasında “dersi bırakma eğilimindeki öğrencilerin tahmini” için bir modül olduğu görülmektedir [4]); ancak daha değişik senaryolar için bu sistemden çekilecek veriler üzerinde çalışan bağımsız uygulamalar geliştirilmesi gerekecektir.

Degree Compass (Austin Peay State Üniv., ABD, 2011) [1]: Bu uygulamada öğrencinin geçmiş başarı durumu ve bulunduğu programın müfredatı dikkate alınarak en yüksek olasılıkla başarılı olabileceği dersler belirlenmekte ve önerilmektedir.

- *AWE: Automated Wellness Engine* (New England Üniv., Avustralya, 2011) [1]: Bu uygulamada öğrencilerin eğitim dönemi boyunca duygusal/fiziksel nasıl hissettiklerine dair periyodik olarak verdikleri geri-bildirim ile çeşitli ders faaliyetleri ve üniversite sistemleri üzerindeki diğer faaliyetleri girdi olarak kullanılmakta ve öğrencinin okulu bırakma olasılığı tahmin edilmeye çalışılmaktadır.
- *İlk yıl başarı kestirimi* (Bucknell Üniv., ABD, 2016) [5, 6]: Bu uygulamada amaç öğrencinin kayıt öncesi ve ilk dönem başarı verisini kullanarak ilk yılındaki başarı ortalamasını tahmin etmek ve düşük başarı potansiyeli olduğu görülen öğrencilere yönlendirme sağlamaktır.
- *University Innovation Alliance (UIA)* [7]: 2014 yılında ABD’deki 11 üniversite bir araya gelerek özellikle kestirimci analitik yaklaşımlarını yenilikçi bir şekilde uygulamayı hedeflemiştir. Bu proje kapsamında aşağıdaki uygulamalar geliştirilmiştir:
- *Zamanında mezuniyet kestirimi* (Uni. of Texas at Austin, ABD): Öğrencilerin demografik ve akademik verileri üzerinde geliştirilen bir model ile zamanında mezun olup olamayacakları tahmin edilmekte ve mezun olamayacak durumdaki öğrencilere çeşitli yönlendirmeler yapılmaktadır.
- *Uyarlamalı öğrenme ve eAdvisor* (Arizona State Üniv., ABD) [8, 9]: Uyarlamalı öğrenme kapsamında ÖYS üzerinden verilen giriş seviye cebir dersi içeriğinin ve hızının her öğrencinin bilgi seviyesine uyarlanması sağlanmıştır. Ayrıca benzer çalışma bir çalışma biyoloji alanında da yapılmıştır [8]. Burada amaç özellikle konu anlatımı aşamasında öğrenciye bilgi birikimine ve anlama kapasitesine göre dinamik olarak içeriği değişebilen dersler sunmaktır. Aynı üniversitedeki eAdvisor sistemi de öğrenciyi ders seçiminde uygun derslere yönlendirmekte ve öğrencinin programdaki başarı durumunu kestirmekte kullanılmaktadır.

Türkiye’de Yükseköğretimde Büyük Veri için Potansiyel Uygulama Alanları ve Yapılabilirlik

Hedef uygulamalar

Önceki kısımda özetlenen örnek uygulamalar da dikkate alınarak, ülkemizde pilot üniversitelerde geliştirilecek büyük veri projelerinde aşağıdaki uygulama hedeflerinden bir veya birkaçına odaklanılabileceği düşünülmektedir:

- Öğrencinin belirli bir derste veya program genelindeki başarısının kestirimi,
- Öğrencinin mezuniyet sonrası istihdam potansiyelinin kestirimi,

-
- Öğrenciler için ders öneri sistemi,
 - Öğrenci değerlendirme formlarındaki serbest metin üzerinde otomatik görüş analizi yöntemleri ile ders/bölüm seviyesinde daha detaylı geri-bildirim elde edilmesi,
 - Fakülte/üniversite genelinde farklı seviyede çok sayıda öğrenciye verilen servis dersleri veya dil derslerinde öğrenci bilgi ve kapasitesine göre uyarlamalı eğitim içeriği sunulması

Bu uygulamalarda bilgisayar biliminin makine öğrenmesi (gözetimli/gözetimsiz öğrenme, vb.), veri madenciliği, öneri sistemleri gibi alanlarından yöntemlerin kullanılması gerekecektir. Ayrıca, kullanılacak veri ölçeği ve türüne göre büyük veri veya bulut bilişim alanındaki yazılım araçlarına da ihtiyaç duyulabilecektir.

Yapılabilirlik ile ilgili noktalar:

- *Hedef uygulama seçimi:* Projenin yapılacağı pilot üniversiteler potansiyel hedef uygulamalar arasından kendi ihtiyaçlarına en uygun ve ellerindeki veri ile gerçekleştirilebilecek olan(lar)ı seçmelidirler. Takdir edileceği üzere bir uygulama için gerekli veri her üniversitede yeterli seviyede bulunmayabilir. Örneğin, “mezuniyet sonrası iş bulma” modelinin eğitilmesi için geçmiş yıllardaki mezunların ilk işini mezuniyetten ne kadar süre sonra bulduğu bilgisi gerekecektir ki, bu bilginin toplanması zordur. Diğer taraftan öğrencilerin bir dersteki/programdaki başarısını tahmin edecek bir model eğitmek için gereken verinin elde edilmesi, eğer o kurumda dersler bir Öğrenme Yönetim Sistemi (Learning Management System) üzerinden yürütülüyorsa, daha kolay olabilir.
- *Donanım, yazılım ve personel altyapısı:* Geliştirilecek projede veri ilgili üniversitenin kullanmakta olduğu dosya sistemi, veri taban(lar)ı ve/veya Öğrenme Yönetim Sistemi üzerinden çekilecektir, buna göre geliştirilecek Büyük Veri uygulaması için gereken donanım/yazılım gereksinimleri değişecektir (ve bir üniversitede geliştirilen uygulama diğerlerinde çalışmayabilecektir). Örneğin, bir üniversitede çalışmakta olan Moodle sistemi üzerine eklenti olarak geliştirilecek bir uygulama daha az ilave kaynak gerektirebilir; ama kendi başına bir uygulama geliştirilecekse üniversitedeki diğer yazılımlarla uyumlu bir altyapı üzerinde geliştirilmesi gerekecektir. Ayrıca yazılım geliştirme, test ve kullanıma alma aşamaları için bilgisayar donanımına ihtiyaç duyulacaktır. Son olarak bahsedilen yazılım geliştirme süreci için tercihen makine öğrenmesi, büyük veri ve veri bilimi alanlarında bilgisi olan yazılım geliştirme personeli gerekmektedir.
- *Yanlılık, veri mahremiyeti ve güvenliği:* Geliştirilecek modellerin yanlılık bakımından hassasiyetinin incelenmesi önemlidir. Örneğin, öğrencilerin bir yüksek lisans programına kabulüne karar vermek için eğitilen bir makine öğrenmesi modelinin, belli bir ideoloji, din, ırk, vb. mensubu kişileri avantajlı (veya dezavantajlı) duruma getirmemesi gerekmektedir. Seçilen hedef uygulamaya göre, model eğitiminde kullanılacak verinin örnekleme ve öznitelik seçimi süreçleri ve/veya model çıktıları yanlılık açısından değerlendirilmelidir. Ayrıca geliştirilecek uygulamalarda veri mahremiyeti/güvenliği konusunda da ilgili standart ve düzenlemelere uyulması önemlidir.

Kaynakça

- [1] Hannover Research, "Learning Analytics for Tracking Student Progress", 2016. <https://www.imperial.edu/docs/research-planning/7932-learning-analytics-for-tracking-student-progress/file>
- [2] Aytacı, Z., Bilge, H. Ş., "Big Data Analytics in Higher Education: A Systematic Review", Internet Applications and Management, 2020. <https://dergipark.org.tr/en/download/article-file/957999>
- [3] Arnold, K. E, Pistilli, M. "Course Signals at Purdue: using learning analytics to increase student success", LAK12, 2012. <https://er.educause.edu/articles/2012/7/signals-using-academic-analytics-to-promote-student-success>
- [4] https://docs.moodle.org/311/en/Students_at_risk_of_dropping_out
- [5] https://digitalcommons.bucknell.edu/cgi/viewcontent.cgi?article=1073&context=fac_conf
- [6] <https://www2.deloitte.com/us/en/insights/industry/public-sector/improving-student-success-in-higher-education.html>
- [7] <https://www.highereddive.com/news/what-a-predictive-analytics-experiment-taught-11-colleges-about-sharing-dat/552986/>
- [8] <https://news.asu.edu/20190820-solutions-asu-develops-world-first-adaptive-learning-biology-degree>
- [9] <https://eadvisor.asu.edu/institutions>
- [10] Mehrabi vd., "A Survey on Bias and Fairness in Machine Learning", ACM CSUR, 54, 6, 2022.

EĞİTİMDE BÜYÜK VERİ KULLANIMI

Trabzon Üniversitesi Görüş Raporu

Günümüzde veri her türlü dijital araçtan elde edilebilmekte ve bulut teknolojileri sayesinde ortak platformlarda depolanabilmektedir. Ancak alışlagelmiş veri tabanı anlayışında bu kadar farklı kaynaktan, farklı formatlarda ve ilişkilendirilmemiş verinin birbiri ile konuşması mümkün değildir. Bu nedenle yakın geçmişe kadar ziyaret edilen siteler, tıklama sayıları, gidilen yerler, konum bilgileri, sağlık ya da öğrenim bilgileri gibi üretilen verilerin oldukça sınırlı şekilde ve sadece kendi bağlamında değerlendirilmekte büyük bir kısmı ise arşiv haline gelmekteydi. Bu durum verinin sadece yığın olarak bağımsız veri tabanlarında saklanmaktan öteye geçememesine neden olmaktadır. Ancak bulut teknolojilerindeki gelişmeler ve veri madenciliği teknikleri verinin oluşumunu, erişilebilirliğini ve birleştirilebilmesini sağlamıştır. Dolayısı ile bu zamana kadar anlamsız görünen veri yığınları artık belirli bir amaç doğrultusunda anlamlandırılabilir hale gelmektedir.

Belirli bir amaç için herhangi bir anda farklı veri kaynaklarından ve farklı formatlarda toplanabilen yeterli miktarda, değerli ve güvenilir, işlenmiş ya da ham veri kümesine büyük veri adı verilmektedir. Büyük verinin 3V olarak da nitelendirilen Hacim (Volume), Çeşitlilik (Variety) ve Hız (Velocity) özelliklerine sahip olması gerekir. Büyük veri kullanımı arttıkça daha kullanılabilir sonuçlar elde edebilmek için bu 3 kavrama Doğruluk (Veracity), Değer (Value) ve Değişkenlik (Variability) özellikleri de eklenmiştir.

Büyük veri, sağlıktan borsaya reklamcılıktan ulaşıma pek çok alanda tahminlerde bulunabilmek için kullanılmaktadır. Bu alanlar içerisinde eğitim, nispeten yeni olmakla birlikte sağlayabileceği avantajlarla büyük bir potansiyele sahiptir.

Büyük veri başta bireysel öğretim olmak üzere eğitimde geniş bir kullanım alanına sahiptir. Çünkü artık öğrenim sadece okul sınırları içinde değildir. Özellikle Covid-19 pandemisi ile uzaktan eğitime geçişin hızlanması, dijital öğrenim verilerinin katlanarak artmasını sağlamıştır. Artık öğrencilerin;

- Arama geçmişleri,
- Ziyaret ettikleri eğitim platformları,
- Ziyaret sıklıkları,
- Bu platformlarda geçirdikleri süreler,
- Tıklama sayıları,
- Web trafikleri,
- Tıkladıkları içerikler,
- Bu içeriklerdeki ilerleme durumları,
- İçerik tercihleri,
- Bitirdikleri kurslar,
- Çeşitli video izleme platformlarında izledikleri videolar,
- Konum verileri
- Alışveriş tercihleri (aldıkları ürünler, alma sıklıkları, markalar, mekanlar, zamanlar vb.)

-
- Kamera görüntüleri ve ses verileri
 - Görünüm ve kıyafet tercihleri
 - Sosyal medyada eriştikleri içerikler,
 - Beğenileri vb. her türlü hareketleri eğitim-öğretim süreçleri için anlamlı bir veriye dönüştürülebilmektedir.
 - Bu veriler beş ana veri türü olarak sınıflandırılabilir.
 - Kişisel veri
 - Öğrencilerin elektronik öğrenme sistemleriyle etkileşimine ilişkin veriler (e-ders kitapları, çevrimiçi kurslar)
 - Öğrenme materyallerinin etkinliğine ilişkin veriler
 - İdari (sistem çapında) veriler
 - Projeksiyon verileri

Önceden birlikte bir anlam ifade etmeyen bu veriler artık bir çıkarım yapabilmek için anlamlandırılabilir. Böylece bu veriler genel olarak öğrencilere;

Öğretim elemanı ve öğrenci açısından:

- Öğrenci ve öğretim elemanlarına eğitim içeriği ve çeşidi önermek,
- Öğrencilerin öğrenim süreçlerini takip ederek başarı durumlarına göre hem öğrenci hem de öğretim elemanına önerilerde bulunmak
- Öğrencilerin yüz ifadeleri ve seslerinden ders süreci ile ilgili karar vermek
- Öğretim elemanına öğretim süreciyle ilgili önerilerde bulunmak
- Ders/bölüm/okul önerileri yapmak,
- Mesleki anlamda yönlendirmek,
- Öğrenme eksiklerinin belirlenmesi,
- Öğretim süreci ile ilgili eksikliklerinin giderilebilmesine yönelik yönlendirmeler yapmak gibi amaçlarla kullanılmaktadır.

Yöneticiler ve karar vericiler açısından

- Yöneticilere öğretim ortamının düzenlenmesi ile ilgili önerilerde bulunmak
- Yöneticilere kampüs içi yerleşim ve çeşitli imkanların (internet erişimi noktaları ve erişim yoğunluğuna göre balans yapılması, otobüs/minibüs güzergah ve duraklarının belirlenmesi, kütüphane kullanım saatlerinin ve personel ihtiyacının belirlenmesi, kantin, yurtlar, öğrenci işleri vb. yerlerin iş yoğunluklarına göre personel ve zaman ayarlamalarının yapılması, güvenlik düzenlemelerinin yapılması, dersliklerin kullanım rutinlerinin ve yoğunluklarının belirlenmesi ve buna göre ısıtma, temizlik, personel, teknik donanım ve ders alanı ihtiyaçlarının düzenlenmesi)
- Düzenlenecek sosyal etkinlikler, şenlikler, söyleşiler, konferanslar vb. etkinlikler için uygun kişileri, uygun yerleri ve zamanları belirlemek
- Düzenlenecek etkinlikler için organizasyon için en uygun zaman, yer, personel ihtiyacı, katılımcı sayısı vb. verileri memnuniyeti sağlayabilecek ve en optimum olacak şekilde belirleyebilir.

Öğrencilerin akademik başarılarının belirlenmesi için büyük veri kullanımı, şimdiye kadarki tek düze sınav, ödev, proje, performans değerlendirmesinin ötesine geçerek derse katılım ve ders için harcanan çabanın da farklı kaynaklardan elde edilecek verilerle ölçülmesine olanak sağlayabilmektedir. Örneğin;

- Belirli bir dersi alan öğrencilerin belirli platformlarda yaptıkları yorumlar, beğeni sayıları, arama seçenekleri, geçirdikleri süre vb. internet verileri incelenerek derse yönelik ilgilerinin ölçülmesi ya da bir eğitmenin performansının belirlenmesi sağlanabilmektedir.

Eğitimde büyük veri, derslerin daha etkin hale getirilebilmesi amacıyla da kullanılmaktadır. Örneğin öğrencilerin;

- Sınavlarda başarısız olduğu konular,
- Hata sayıları,
- Çözüm adımı sayıları
- Harcadıkları süre verileri eğitimcilerin ders için paylaştıkları içeriklerin yeterliklerinin belirlenmesinde kullanılabilir.
- Öğrencilerden alınacak ayrıntılı girdiler ile içeriğin yeterliliğinin belirlenmesi ve eksiklerinin giderilmesi ile en uygun içeriklerin oluşturulması ve öğretim tasarımının iyileştirilmesi sağlanabilmektedir.
- Çeşitli alışveriş ve video akış sitelerinin müşterilerine öneri amaçlı yaptığı çıkarımlar öğrencilerin web hareketlerinin izlenmesi yoluyla uygun eğitim içeriklerinin getirilmesinde kullanılmaktadır.
- Bireysel öğretimi sağlayacak uyarlanabilir sistemler ve zeki öğretim sistemleri için analizler gerçekleştirilebilir.

Eğitim-öğretim ile ilgili süreçleri yalnızca dersler boyutunda düşünmememiz gerekmektedir.

- Örneğin öğrencilerin web geçmişlerinden hareketle ilgi alanlarının belirlenerek onlara yönelik sosyal, sportif ya da sanatsal etkinlikler düzenlenebilmektedir. Aynı durum bireysel olarak öğrenciler için ilgi alanlarına uygun faaliyetlere yönlendirmelerinde kullanılabilir.
- Web hareketi verisine, öğrencinin akademik başarısını, daha önce ziyaret ettiği yerleri, önceki başarılarını, deneyimini ve ilgi alanlarını da ekleyerek üniversite ve bölüm tercihi ya da sonraki aşamada meslek tercihi için çıkarımlarda bulunulabilmektedir.
- Üniversiteler, öğrencilerin kampüse giriş çıkışlarını ve çeşitli etkinliklerini kimlik kartları yardımıyla takip etmektedir. Bir öğrencinin kampüs içi hareketi azalıyorsa, üniversite personeli nedeni belirleyebilir ve yardım önerebilir.
- Okul terk oranının düşürülmesi için öğrencinin hareketleri, ders başarıları vb. verileri üzerinden değerlendirmeler yapılabilir

Eğitimde büyük veri kullanımı sadece öğrenciler değil kurumlar için de yaygınlaşmaktadır.

- Örneğin eğitim kurumları daha gerçekçi ve tarafsız sonuçlar elde edebilmek amacıyla öğrenci ve personel memnuniyetlerini anketler yerine sosyal medya ve web platformlarından gelen yorumları analiz ederek belirleyebilmektedir.
- Yine kurumlar verimlilik içinde büyük veriyi kullanabilmektedir. Örneğin kampüsteki ağ günlüklerini izleyerek ağ trafiğini düzenleme ya da öğrenci tercihlerini analiz ederek verilen hizmetle ilgili tüm bileşenlerin yeterliklerinin belirlenmesi sağlanabilir. Yine kampüs içi

verilerden öğrencilerin ihtiyaçları belirlenebilir.

Eğitimde büyük veri çalışmalarının görülmeye başlandığı 2006 yılından bu yana çalışma sayısında katlanarak artan bir ivme görülmektedir. Özellikle 2014 yılından günümüze yoğunlaşan büyük veri çalışmalarının; eğitim-öğretim sürecinde yer alan aktörlerin üretebildiği veri türlerinin çeşitlenmesi ve veri üretme ve paylaşma hızlarının artmasıyla daha kapsamlı ve yaygın olacağı aşikârdır. Bu çalışmalardan elde edilecek verilerin ise özellikle bu süreçlerdeki eksikleri, ihtiyaçları ve yapılması gerekenleri belirlemek açısından önemi daha iyi anlaşılacaktır. Artık sadece insana özgü tecrübeyle elimizdeki verileri kullanarak karar verme süreci yerini geçmişten günümüze gelen ve hatta o anda oluşan çok farklı kaynaklardan verileri değerlendirerek karar verebilen sistemlere bırakmaya başlamıştır. Tüm alanlarda kendini gösteren bu durum eğitim öğretim süreçleri içinde kaçınılmaz olarak önümüzdeki yıllarda büyük veri kavramından daha çok bahsedilmesine neden olacaktır.

Büyük Veri Projesi (BVP) Yapılabilecekler

“Büyük veri projesi (BVP)” kapsamında üniversitelerin dijital verilerinin ortak toplanacağı bir veri havuzu oluşturulabilir.

Bunun için öncelikle üniversitelerin kullandığı sistemler için asgari formatların belirlenmesi gerekmektedir.

- Uzaktan eğitim sistemleri (LMS ve sanal sınıf), öğrenci kampüs kartı verileri, kütüphane bilgileri, personel bilgileri, ders kayıtları, kamera kayıtları, sınav notları, öğrencilerin kulüp ve sosyal etkinlik verileri, web sitesi erişim bilgileri, yemekhane, yurt, internet kullanım verileri, sosyal alanlar vb. internet erişim noktaları ve ortak ağlardan elde edilen veriler, üniversitelerin kullandığı otomasyon sistemlerinin bir listesi ve veri tabanı ER diyagramları istenerek bunlardan alınabilecek veriler belirlenir.
- Sağlanabilecek veriler için etik kurallar ve KVKK kapsamındaki yasal zorunluluklar için gerekli tedbirler alınır.
- Bu verilerin toplanabileceği bir sunucu hizmeti sağlanır
- Verilerin sunucuya düzenli aktarımı için servisler yazılır
- Üniversitelerde bu konuda sorumlu personel görevlendirmeleri yapılır
- Üniversitelerle BVP arasındaki entegrasyon sağlanır
- BVP sunucu üzerinde analitik yazılımları koşularak
 - o İstenen içgörüler elde edilir,
 - o Öngörüler oluşturulur,
 - o Karar destek sistemleriyle YÖK tarafından alınacak kararlar için sistemden görüşler elde edilebilir
 - o Üniversiteler için tüm kaynakların optimizasyonu sağlanır
- BVP sunucusu üzerinde Makine Öğrenmesi algoritmaları koşularak
 - o Sistemin gelen verilerden öğrenmesi ve hem YÖK’ün ilgili komisyonları hem üniversiteler hem de öğretim elemanları için dijital asistan hizmetinin verilmesi sağlanabilir. Dijital asistan uygun veriler sağlandığı sürece;
- YÖK için üniversitelere kaynak tahsisi, üniversitelerin değerlendirilmesi, akreditasyon süreçleri ve Kalite Komisyonlarına destek vb. amaçlarla kullanılabilir.

-
- Üniversite yönetimleri için tüm karar süreçlerinde danışmanlık yapabilir.
 - Öğretim elemanları için derslerine katılım oranları, öğrenci görüşleri ile içeriklerin ve ders süreçlerinin değerlendirilmesi, önerilerde bulunulması sağlanabilir.
 - Derin öğrenme ile kamera ve ses verileri işlenerek öğrenci kampüs kullanımı, başarıyı etkileyen etmenler, memnuniyet düzeyi, okul bırakma nedenleri, talepler vb. içgörüler elde edilebilir.

Kısa, orta ve uzun vadeli iş-zaman takvimleri ve ilgili iş paketleri için alt teknik ekipler oluşturularak süreç planlanabilir. Aşağıda yapılması planlanan işler ve süreleri ile ilgili genel bir süreç planı sunulmaktadır. Belirtilen işlemler alt iş paketlerinden oluşmaktadır ve her bir iş paketi kendi içinde detaylandırılmalıdır. Proje iş paketlerinin detaylandırılması, görev dağılımının ve koordinasyonun belirlenmesi süreçleri proje öncesi asgari 3 ay içerisinde sağlanabilir.

- Öncelikle üniversite sistemlerinin detaylı analizi ve sağlayabilecekleri verilerin belirlenmesi ile süreç başlatılır (ilk 6 ay). Sunucu kurulumları (ilk 6 ay), servislerin yazılması (ilk 9 ay) ve pilot üniversitelerin entegrasyonları (ilk 12 ay içinde) sağlandıktan sonra projenin;
 - o 1. Aşamasında YÖK ile üniversiteler arasında büyük veri işleme süreçleri tamamlanır (Projenin 24. ayına kadar). Bu kapsamda YÖK'ün üniversiteler ile ilgili verileri işleyerek kendi içgörülerini oluşturması sağlanabilir.
 - o 2. Aşamasında BVP sunucusu üzerinde oluşturulacak dijital asistan ile Üniversite yönetimlerine mevcut durum analizleri ve öneriler sunulur (Bu iş paketi kendi içinde detaylandırılmalıdır. Pilot üniversiteler için projenin 36. ayına kadar tamamlanabilir).
 - o 3. Aşamasında BVP, öğretim elemanlarının erişimine açılarak kullanımı sağlanabilir (Bu iş paketi kendi içinde detaylandırılmalıdır. Pilot üniversiteler için projenin 48. ayına kadar tamamlanabilir).

YÜKSEKÖĞRETİMDE BÜYÜK VERİNİN KULLANIMINA İLİŞKİN İSTANBUL TOPLANTISINDA SUNULAN GÖRÜŞLER



Önerilen Uygulama Alanları

- Eğitim kalitesini takibi,
- Mezun istihdam takibi,
- Akademik üretkenliğin iyileştirilmesi,
- Yayın performansının takibi,
- Araştırmaların ve atıfların etki değeri gücü takipleri,
- Üniversitelerin akademik performans açısından değerlendirilmesi,
- Verilerin stabil olması gerektiği, veri mahremiyeti çalışması,
- Verilerin nasıl elde edileceği, hangi sistemler üzerinden sağlanacağı,
- Verilerin kullanımında ortaya çıkabilecek güçlükler göz önünde bulundurulmalıdır (yasalar, kullanım izni, verinin kullanımında karşılaşılan zorluk),
- Büyük veri için alt yapı sistemlerinin kurulması (bilginin saklanması, kullanılması).

Pilot Üniversitelerin Seçiminde Önerilen Kriterler

- Pilot üniversiteler içinde araştırma üniversitelerinden bulunmalıdır.
- Seçilecek üniversitelerin dijital ofisleri kurulmuş olmalıdır.
- Büyük Veri Projesi kapsamında, pilot üniversitelerin alanla ilgili iş güçleri bir araya getirilebilir.
- Pilot üniversitelerde, idari karar destek sistemleri kurulmalıdır.

**BÜYÜK VERİ PROJESİ
ÇALIŞMA GRUBU MAKALELERİ**



YÜKSEKÖĞRETİM EKOSİSTEMİNDE DİJİTALLEŞME VE BÜYÜK VERİ

*Dr. Ali Taha KOÇ

**Doç. Dr. İhsan Tolga Medeni

Dijital dönüşüm kavramı temelde, dijital teknolojiler kullanılarak mevcut iş süreçlerinin iyileştirilmesi veya yeni iş süreçleri oluşturmak suretiyle üretkenliğin artırılmasını ifade etmektedir. Dijital teknolojilerin bugün ulaştığı noktada artık fiziksel dünyada yapılabilen birçok aktivite dijital ortama uyarlanmış, bir anlamda yeni dijital dünyalar kurulmaya başlanmıştır. Eğitim dâhil tüm sektörlerdeki hizmet modelleri, değişen gereksinimler ve fırsatlar dikkate alınarak yeniden şekillendirilmektedir. Hızla gelişen dijital bilgi teknolojilerinin ana itici güç olduğu bu geçiş sürecinde, analog döneme ait geleneksel kavramların ve kurumsal yapıların da dönüşerek dijital dünyada yeni işlevler kazandığı görülmektedir.

Sosyoekonomik yaşamın her alanında hızlanarak ilerleyen dijital dönüşüm sürecinin ortaya çıkardığı en önemli sonuçlardan biri de muazzam ölçekte dijital veri üretilmesidir. Bu süreçte, üretilen verinin hacminin, çeşitliliğinin ve üretim hızının inanılmaz ölçüde artması, “büyük veri” kavramının ortaya çıkmasını sağlamıştır. Bilişim sistemlerinin hizmet süreçlerini hızlandırmasının ötesinde, büyük verinin işlenmesi suretiyle karar mekanizmalarında etkinliğinin önemli ölçüde artırılacağı da gözlemlenmiştir. Bu durum, organizasyonların “veri temelli” olacak şekilde yeniden yapılandırılmasını gündeme getirmiştir. Veri temelli organizasyonlar belirli bir sektöre özel yapılar olmayıp dijital verinin üretildiği tüm sektörlerde ortaya çıkmaktadır. Örneğin, tarımsal faaliyetlerde verimi artırmak üzere toprak kalitesi, hava durumu ve bitki sağlığıyla ilgili verileri sürekli olarak toplayıp tarım politikasını geliştirirken kullanan bir kamu kurumu veri temelli organizasyondur. Hemen her sektör için benzer örnekleri verilebilecek bu tür organizasyonlar, geleneksel yapılara göre üretkenliği ve sosyal refahı önemli ölçüde artırma potansiyeli barındırmaktadır.

Büyük verinin işlenerek anlamlı hale getirilip iş ve karar süreçlerini besleyebilmesi insanların bilişsel kabiliyetlerini fersah fersah aşmaktadır. Bu nedenle, büyük verinin işlenerek veriden bilgiye, iç görüye ve öngörüye ulaşılabilmesi için büyük veri analitiği ve yapay zekâ teknolojileri ön plana çıkmaktadır.

Büyük verinin toplumsal ve ekonomik anlamda çok önemli kazanımlar sağlayabileceği alanlardan biri de yükseköğretimdir. Yükseköğretim ekosistemi bir bütün olarak dikkate alındığında; öğrencilerin kabiliyetlerine ve ilgi alanlarına uygun programlara yönlendirilmesi, araştırma faaliyetlerinin önceliklendirilmesi, akademik başarının ölçülmesi, eğitim içeriklerinin kişiye özel şekilde tasarlanması, iş hayatında ihtiyaç duyulan güncel yetkinliklerin belirlenmesi ile eğitim programları ve akademik personel yetkinliklerinin iyileştirilmesine kadar büyük veriden ciddi anlamda fayda sağlamak mümkündür.

Büyük veriye dayalı karar mekanizmalarının ana girdisinin dijital veri olduğu dikkate alındığında, yükseköğretimde bu potansiyelin ortaya çıkarılmasında ekosistemin tüm unsurlarına yönelik veri havuzlarının oluşturulması ve bunların birlikte analiz edilmesinin önemi açıktır. Türkiye’de hâlihazırda Millî Eğitim Bakanlığı (MEB), Öğrenci Seçme ve Yerleştirme Merkezi, Yükseköğretim Kurulu ve üniversitelerimizde öğrenci, mezun ve akademisyenler için kapsamlı veri havuzları bulunmakta, bu aktörlerin yükseköğretim ekosistemindeki çeşitli faaliyetlerine ilişkin veriler düzenli olarak kayıt altına alınmaktadır.

Ülkemiz adına değerli bir “dijital varlık” olarak kabul edilebilecek bu veri havuzlarından mümkün olan en yüksek faydanın sağlanabilmesi bu verilerin bir araya getirilerek analiz edilmesini gerektirmektedir.

*Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanı

** Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Uzmanı

Örneğin; öğrencilerin yükseköğrenim öncesi akademik başarıları ve ilgi alanlarına ilişkin MEB'in elindeki veriler ile yine öğrencilerin yükseköğrenim sürecindeki faaliyet ve başarılarına ilişkin veriler birlikte analiz edilerek öğrencilerin başarı potansiyelinin daha yüksek olduğu uzmanlık alanlarına yönlendirilmesi, yetkinlik eksikliklerinin ve ihtiyaçlarının daha doğru şekilde tespit edilmesi mümkündür. Böylece hem öğrencilerin bireysel olarak akademik yaşamlarına ilişkin daha isabetli kararlar alması sağlanacak hem de üniversite yönetimleri ve genel olarak politika yapıcılar bu bireysel bilgileri kitle bilgisine, karar ve hedeflere dönüştürüp stratejik hedefler ile uyumunu temin edebilecektir.

Son dönemlerde özellikle pandemi nedeniyle çevrim içi eğitim araçları eğitimde çokça kullanılmaya başlanmıştır. Hem eğitim ortamının hem de eğitim içeriğinin dijitalleştiği dikkate alındığında, yükseköğretim kurumlarının bu dönüşüme başarılı şekilde uyum sağladığı görülmektedir. Diğer taraftan, öğrenciler farklı platformlardan eğitimler alarak geleneksel eğitim içeriğinin ötesinde bir bilgiye erişim sağlamaktadır. Dijitalleşmenin en önemli unsuru bir ortam ve içerik yaratmak olduğu kadar bu kaynakların kullanılması esnasında üretilecek verilere de odaklanmaktır. Bu nedenle çevrim içi araçlar hem akademik personelin hem de öğrencilerin performanslarının değerlendirilmesinde ve geliştirilmesinde kullanılacak büyük verinin üretilmesinde asli kaynak olacaktır. Teknolojinin gelişimiyle sınav sistemi, eğitim süresi ve içeriğe erişim gibi konularda çok daha kapsayıcı ve iç içe geçmiş bir yapısal dönüşüm söz konusu olacaktır.

Yukarıda açıklanan mekanizmayı bir adım daha öteye taşırsak, güncel bilgilerin (nitelikli işgücü açığı olan meslekler, yeni gelişen uzmanlık alanları, mesleklerdeki dönüşümün ortaya çıkardığı yeni yetkinlikler vb.) derlendiği veri tabanları oluşturulup bunların yükseköğretim ekosistemindeki veri havuzlarıyla birleştirilmesiyle; mevcut programların müfredatının iyileştirilmesi, yeni eğitim programları oluşturulması, öğrencilerin araştırma projelerinin ve tezlerinin güncel ve ihtiyaç duyulan alanlara odaklandırılması, burs ve desteklerin bu doğrultuda yapılandırılması gibi yükseköğretim sistemiyle iş hayatını daha sıkı şekilde birbirine eklemleyip ilişkilendirecek mekanizmalar da kurgulanabilir.

Verinin en temel özelliği, analize konu edilecek birbiriyle ilişkili verilerin hacmi ve çeşitliliği arttıkça analizden elde edilecek faydanın da hızla artmasıdır. Bu fayda, var olduğu bilinen bir örüntüye dayalı öngörünün isabet oranını artırmak şeklinde olabileceği gibi olgular arasında daha önce var olduğu bilinmeyen yeni örüntü ve ilişkilerin keşfedilmesi şeklinde de olabilir. Bu nedenle, yükseköğretim ekosisteminde büyük veriden daha etkin şekilde faydalanabilmek için hem bu ekosistemdeki hem de bu ekosistemle yakın ilişkisi bulunan diğer ekosistemlerdeki (ortaöğretim ve iş hayatı gibi) olay ve aktörlerle ilgili verilerin hacim ve çeşitliliğini, büyük ölçekte veri toplama ve işlemenin getirdiği teknik altyapı maliyeti de dikkate alınarak, mümkün olduğunca artırmak gereklidir. Diğer bir ifadeyle, yükseköğretim ekosistemi ve bununla ilgili diğer ekosistemlerdeki kurumların "veri temelli organizasyon" niteliği geliştirilmelidir.

Elbette, muhtemelen farklı aktörler ve kurumlar tarafından oluşturulacak bu ayrık veri havuzlarının da analiz edilebilir şekilde bir araya getirilmesi ihtiyacı olacaktır. Diğer taraftan, veriden anlam çıkarmak ilgili veri kümelerinin bağlamsal doğası hakkında da bilgi sahibi olunmasını gerektirmektedir. Dolayısıyla, farklı kurum ve kuruluşların elindeki yükseköğretimle ilgili ayrık veri kümelerinin birlikte analiz edilmesine imkân sağlayacak ve bunun için gerekli farklı alan uzmanlıklarını bir araya getirebilecek yönetim ve koordinasyon mekanizmalarına da ihtiyaç olacaktır. Aslında analog dönemde iş ve işlemlerin etkin yürütülmesi için ayrışan kurumlar, dijital dönemde veri temelli olarak yeniden kaynaşmaktadır. Bu ihtiyaç doğrultusunda kurulan Dijital Dönüşüm Ofisi; veri temelli dönüşümün sağlanmasına yönelik bulut bilişim stratejisi, ulusal veri sözlüğü, kamu veri alanı, siber güvenlik rehberliği ve e-devlet kapısı (Dijital Türkiye Platformu) entegrasyonları gibi önemli çalışmalar yürütmektedir.

Yükseköğretim alanında kurulacak bir entegre veri yönetimi mekanizmasının hem öğrenci ve ailelerin yükseköğretimden beklentilerini doğru şekillendirmek ve gençleri doğru yönlendirerek başarılı ve mutlu olacakları meslekler edinmelerine yardımcı olmak hem de ülkemizin farklı alanlarda ihtiyaç duyduğu nitelikli insan kaynağını daha hızlı ve etkin şekilde yetiştirmek açısından çok önemli bir potansiyel barındırmaktadır.

Bireyler, yükseköğretim ekosistemine tekil kimlik numarası ile dâhil olmaktadır. Dolayısıyla veri temelli organizasyonel dönüşüm için uçtan uca bir profil oluşturma ve kişiselleştirme için temel parametre hazırır. Kişiselleştirme mekanizması, öğrenciler için ilkokuldan üniversiteyi bitirip, iş hayatında bir mesleğe yerleşene kadar takip edilebilecek adımlar silsilesi şekliyle ortaya konulabilir. Öğrencinin yükseköğretim öncesindeki ve sırasındaki akademik başarıları ile sosyal hayatında oluşturduğu veriler dikkate alınarak ve işgücü piyasasına ilişkin güncel bilgi ve öngörülerle harmanlanarak, oluşturulmuş profiline göre öncelikli mesleklere veya uzmanlık alanlarına yönlendirilmesi sağlanabilir. Bu sayede öğrencinin gerçekten kendi becerilerine ve kişiliğine uygun eğitimi alıp, ömrü boyunca gerçekten de ait olduğunu hissettiği ve başarılı olacağı mesleği edinmesi sağlanabilir. Diğer taraftan, yapay zekâ ve veri analitiği teknolojileriyle oluşturulacak bu tür büyük veri sistemlerinin, kullanılan veri kümelerindeki yanlışlık ve eksiklikler nedeniyle her zaman doğru sonuç ve öngörüler üretmeyebileceği dikkate alınmalı ve bu tür mekanizmalar sadece bir karar destek sistemi olarak tasarlanmalıdır.

Türkiye'nin yapay zekâ alanındaki ilk ulusal strateji belgesi olan ve Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ile Sanayi ve Teknoloji Bakanlığı tarafından hazırlanıp 2021 yılı Ağustos ayında uygulamaya konan Ulusal Yapay Zekâ Stratejisi'nde (UYZS), yükseköğretim de dâhil birçok alanda yapay zekâ teknolojilerinden daha etkin şekilde faydalanmaya yönelik tedbirlere yer verilmiştir. Bu tedbirler, ülkemizin ihtiyaç duyacağı nitelik ve nicelikte yapay zekâ uzmanlarının yetiştirilmesi, farklı kurumların elindeki veri kümelerinin büyük veri analizi amacıyla bir araya getirilmesi gibi çalışmaları kapsamaktadır. Dolayısıyla, UYZS kapsamında yürütülmesi öngörülen bu çalışmalarla, ülkemizdeki yükseköğretim ekosisteminin yukarıda açıklandığı şekilde "veri temelli organizasyon" niteliğinin güçlendirilmesi ve büyük veriden daha fazla faydalanmasına katkı sağlanacaktır.

YÜKSEK ÖĞRETİMDE VERİ YÖNETİŞİMİ

*Prof. Dr. Erkay Savaş

Özet: Bu yazımızda bir eğitimci, akademik yönetici ve bir bilgisayar mühendisi olarak yüksek öğretim veri yönetişimi konusundaki kişisel düşüncelerimizi, tecrübelerimizi ve ileriye dönük öngörülerimizi paylaşacağız. Yazımıza, doğru veri yönetişiminin bir yüksek öğretim kurumu için neden gerekli olduğunu ve hangi faydaları sağlayabileceğini belirterek başlayacağız ve dönüştürücü etkisine de vurgu yapacağız. Değineceğimiz konuların arasında, verilerin toplanması; doğruluğunun teyit edilmesi; güvenlik ve güvenilirliklerinin sağlanması; mahremiyetlerinin korunması; kolay, hızlı ve çok noktadan erişimin sağlanması; erişim yetkilendirme şemalarının oluşturulması ve uygulanması; raporlama ve iş zekası olarak da bilinen veri işleme ve anlamlandırma yazılım araçlarının etkin kullanılması yer alacaktır. Veri yönetişimi için yukarıda saydığımız konuların doğru adreslenebilmesinin önünde önemli teknik sorunların yanı sıra, idari, hukuki ve davranışsal zorluklar da yer almaktadır.

Giriş

Yüksek öğretim kurumları günlük operasyonel faaliyetleri için, bilgi işlem teknolojilerinden her geçen gün daha çok faydalanmaktadır ve bu teknolojiler büyük hacimlerde ve çeşitlilikte veri üretmektedir. Son zamanlarda tüm dünyanın gündemini meşgul eden COVID-19 pandemisi, tüm dünyada yüksek öğretimde meydana gelen dönüşüm hareketlerini hızlandırıcı bir etki yaratmıştır. Pandemi koşullarında önce eş-zamanlı çevrim içi, aşı bulunmasından sonra ise eş-zamanlı hibrit eğitim yöntemleri tüm üniversiteler tarafından, biraz da zorunluluktan uygulanmaya başlanmıştır. Eş-zamanlı eğitimin yanı sıra, yüksek kaliteli ders malzemelerinin (sunum, görsel içerik ve ders videoları) hazırlanıp öğrenciye ders öncesi sunulduğu ve öğrenciden yüzyüze yapılan derse hazırlıklı gelmesinin gerekli olduğu ve aktif öğrenmeyi teşvik eden yeni, ileri eğitim teknikleri de daha çok uygulanmaya başlamıştır. Önce de belirttiğimiz gibi, pandemi, arka planda güçlü bir şekilde gelişmekte olan bu değişimi, sadece hızlandırmış ve yüksek öğrenimin yakın gelecekte evrileceği biçim hakkında bir fikir oluşmasını sağlamıştır. Konuya daha kapsayıcı bakanlar için ise, bu dönüşümün getireceği fayda ve yeni olanaklar da daha belirgin bir hale gelmiştir. Pandemi sırasında uyguladığımız yeni eğitim yöntemlerinin bir kısmının süreklilik kazanacağına kesin gözüyle bakılmaktadır.

Bu dönüşümün en önemli unsurlarından bir tanesi de derslerin işlenmesinde, öğrenme yönetim sistemleri (learning management systems) benzeri yazılım araçlarının ve bu yazılım araçlarının ileri özelliklerinin daha çok kullanılmasıdır.

Bilgisayar Bilimleri ve Mühendisliği Öğretim Üyesi

Bunun neticesinde, öğrencinin eğitime katılma biçimi, çalışma alışkanlıkları ve başarımı hakkında daha fazla veri üretilebilmektedir. Örneğin, çevrim-içi ders içeriklerine ne zaman ve ne sıklıkta erişildiği, öğrencinin verilen kısa çevrim-içi sınavlarda hangi sorularda ne kadar zaman harcadığı gibi veriler eğitimcinin ders yönetişimini iyileştirmek için çok önemli bilgiler içermektedir.

Bunun dışında, pandemi öncesinde de halihazırda toplanan ve öğrenci, eğitim, idari ve akademik başarımını ölçmede kullanılacak veriler de yüksek öğretim kurumunun dönüşümünün hızlandırılmasında, stratejik hedeflerinin belirlenmesinde ve takip edilmesinde kullanılabilir. Bir yüksek öğretim kurumunun stratejik hedeflerine ulaşabilmesi için belirli bir plan ve takvim çerçevesinde bir takım faaliyetleri yürütmesi ve bir takım anahtar başarım ölçütleri için gerekli verileri toplaması ve işlemesi gerekmektedir. Örneğin, öğrenci, akademisyen ve idari çalışanlarının başarım ölçüm ve takibinde uygun

*Sabancı Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi Dekanı

ve elde edilebilir başarıml ölçütleri seçilmeli, toplanmalı ve Üniversitenin stratejik hedeflerine erişmesine katkısı için bir sistematik geliştirilmelidir.

Adreslenmesi Gereken Konular

Yukarıda bahsi geçen veriler kurumun içerisindeki farklı birimler tarafından üretilmekte ve/veya toplanmaktadır. Birimler bu verilerin paylaşılmasında, güvenlik ve mahremiyet unsurlarını da düşünerek ve çoğu zaman da haklı olarak, isteksiz davranmaktadır. Ayrıca, çalışanların güncelliğini yitirmiş teknoloji ve süreçleri kullanmakta ısrar ettikleri zaman zaman gözlemlenmektedir. Örneğin, daha güncel veri tabanı yazılım teknolojileri uzun zamandan bu yana var olmalarına rağmen, birimler arası veri paylaşımı tablolama yazılımları kullanılarak yapılmaya devam etmektedir. Veri paylaşılmadığında ya da bu şekilde paylaşıldığında, toplanan diğer tür verilerle birleştirilememekte ve veriden çıkartılabilecek faydalı bilgilerin nitelik ve niceliği bundan olumsuz yönde etkilenmektedir. Ayrıca, gereksiz iş gücü ve zaman kaybı da not edilmelidir. Üretilen bu verilerin büyük veri niteliği taşıdığını ve bununla bağlantılı teknik zorlukları da düşünürsek, bahsedilen dönüşümün kolay olmayacağı anlaşılabilir.

Herşeyden önce verinin doğru bir şekilde toplanması; doğruluğunun ve erişilebilirliğinin sağlanması ve çeşitli risklere karşı korunması gereklidir. Verilerin güvenliğinin sağlanması ve mahremiyetinin korunması da önemli unsurlar olarak öne çıkmaktadır. Toplanan ve işlenen veriler, önemli ölçüde kişisel bilgiler içerdiğinden Kişisel Verileri Koruma Kanunu'na¹ (KVKK) göre saklanmalıdır. Bununla beraber, operasyonel gereklilikler nedeniyle bu verilerin işlenmesi gerekmektedir ve KVKK'ya göre gerekli önlemler alındığı sürece, kurum içerisindeki yetkili kişilerin bu verileri işleminde bir sakınca yoktur.

Diğer bir konu, ise farklı birimlerden gelen verilerin, operasyonel veri depo (İngilizce, operational data store) yazılımları ile biraraya getirilmesidir. Bu yöntem kullanılarak, veriler operasyonel ihtiyaçlar için yetkili çalışanların hızlı bir şekilde erişimine ve bu veriler üzerinde raporlama vb işlemler yapabilmesine olanak tanınacaktır. Raporlama ya da daha güncel bir terminoloji kullanarak iş zekası yazılımlarının veri işlem ve anlamlandırmada önemli bir rolü vardır.

En son ancak en önemli konulardan bir tanesi de, çalışanların yeni süreçleri ve yaklaşımları benimsemesi için bir farkındalık yaratılmalı, ilgili güncel teknolojileri kullanabilmesi için bir eğitim programı oluşturulmalıdır.

Sonuç ve Öneriler

Yüksek öğretim kurumlarının veri yönetişimini etkin ve verimli gerçekleştirebilmesi için bir takım süreçlerin ve mekanizmaların oluşturulması gerekmektedir. Bunun için bizim önerimiz, dijital dönüşümü yönetecek kurum çapında bir kurulun/komitenin/ofisin oluşturulması ve burada üniversitenin değişik birimlerinin temsil edilmesi yönünde olacaktır. Dijital dönüşüm birimi, kurumda operasyonel ihtiyaçlar için toplanan tüm verilerin envanterini çıkartmalı ve bu envanterde bu verilerin hangi amaçla ve kimler tarafından kullanıldığı belirtilmelidir. Bunun dışında, halihazırda toplanmayan, fakat operasyonel ihtiyaçlar ya da stratejik hedefler için toplanması gereken başarıml ölçütleri ile ilgili verilerin toplanmaya başlanması için bir plan yapılmalıdır. Toplanacak ve işlenecek bu verilerin doğruluğunun ve erişilebilirliğinin sağlanması için gerekli süreçler tanımlanmalı, bu süreçlerin sorumluları belirlenmeli, sürdürülebilir ve ölçeklenebilir bir yazılım mimarisi oluşturulmalı ve bunlara uygun yazılım araçları belirlenmelidir. Büyük hacimleri, çeşitlilikleri ve yüksek üretim hızları nedeniyle toplanan veriler, Büyük Veri sınıfında düşünülmeli ve ilgili Büyük Veri işleme yöntem, teknoloji ve yazılım çözümleri tercih edilmelidir. Seçilen yazılım çözümlerinin açık kaynaklı olması ve dünyada yeterince desteğinin ve genişçe kullanılabilir olması bir tercih sebebi olabilir. Kurumun kapasitesine göre, yazılım çözümlerinin en azından bir kısmının iç kaynak ve personel ile geliştirilmesi tavsiye edilir.

Verilerin güvenlik ve mahremiyetlerini korumak için, KVKK'nın yanısıra, Cumhurbaşkanlığı Dönüşüm Ofisinin genel ve rehberleri² kullanılabilir. Burada, kurumun varlıklarının ve kişisel verilerinin kritiklik

¹ <https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=6698&MevzuatTur=1&MevzuatTertip=5>

seviyesi belirlenmeli ve gerekli güvenlik tedbirleri uygulamaya alınmalıdır. Bu çalışmalar hem teknik bilgi hem de tecrübe gerektirdiği için danışmanlık alınması gerekebilir. Alınacak tedbirlerin mali boyutunun olacağı göz önüne alınarak gerekli finansal planlamaların ve önceliklendirmenin yapılarak bir uygulama takvimi oluşturulmalıdır.

Değişim ve dönüşümün etkin bir şekilde uygulamaya geçirilmesinde insan unsuru en önemli rolü oynamaktadır. Dönüşümün gerekliliği konusundaki iradenin ortaya konulması, farkındalığın arttırılması ve uygun eğitimlerin verilmesi gerekmektedir.

Son olarak, dünyada ve ülkemizde diğer yüksek öğretim kurumlarında veri yönetiřimi konusunda uygulanan en iyi pratikler mutlaka incelenmeli ve kurumun yapı ve ihtiyaçlarına göre uyarlanmalıdır. Kurumlar arası bilgi, tecrübe ve birikim paylaşımı da tüm kurumlar için topyekün bir fayda oluşturacaktır.

² <https://cbddo.gov.tr/bgrehber>

BİLGİ ÇAĞINDA BÜYÜK VERİNİN ÖNEMİ

*Doç. Dr. Ahmet Cüneyd Tantuğ

*Prof. Dr. Şule Gündüz Öğüdücü

Dünya tarihi boyunca teknolojiye bağlı olarak çağlar kapanıp yerine yeni çağlar başlamıştır. Rönesans Çağı'ndan sonra 1750'lerde başlayan Endüstri Çağı (Industrial Age), elektronik ve sayısal teknolojilerdeki gelişmelere paralel olarak 1950'lerde yerini Bilgi Çağı'na (Information Age) bırakmıştır. Su, buhar ve elektrik gücünün kullanımını esas alan Endüstri Çağı gibi önceki çağlarda da fiziksel materyallerdeki buluşlar ve farklı enerji dönüşüm türlerinin kullanılmaya başlanması ana kırılma noktalarını oluştururken, ilk kez Bilgi Çağı'nda ise bu özellik değişmiştir. Bilgi Çağı'nda, elle tutulmayan ve gözle görülmeyen soyut bir kavram olan bilginin saklanması, iletilmesi ve işlenmesi gibi nitelikler kırılma noktalarını oluşturmaktadır. Paleolitik Çağ'dan bu yana çağları kapatıp yeni çağları başlatan fiziksel ve somut teknoloji değişimlerinin yerini soyut teknolojilere bırakması bir paradigma değişikliği olarak nitelendirilebilir ve tek başına bu özellik dâhi Bilgi Çağı'nın öncekilerden ne kadar farklı ve önemli olduğu göstermektedir.

Henüz yaklaşık 70 yıllık bir geçmişe sahip Bilgi Çağı'ndaki gelişmeler ve buluşlara odaklanıldığında, önceki çağlara göre çok daha hızlı ve büyük ölçekli değişimler doğurduğu ve çok daha fazla insanın hayatını etkilediği görülmektedir. Örneğin otomobillerin 50 milyon kişi tarafından kullanılması 62 yıl sürmüştür, cep telefonlarının aynı sayıda kullanıcıya ulaşması 12 yıl, Pokémon GO isimli oyunun ise yine aynı sayıda kullanıcıya ulaşması sadece 19 gün sürmüştür. Çok daha geniş kitlelere çok daha kısa sürelerde ulaşan teknolojik gelişmelerin yaşandığı Bilgi Çağı, kişisel bilgisayarların pazara çıkması, internetin doğuşu, mobil iletişim, sosyal medyanın ortaya çıkması, akıllı telefonların geliştirilmesi, yapay zekâ teknolojilerinin uygulanabilir hâle gelmesi gibi önemli kilometre taşları ile bezelidir.

Bilgi Çağı'nın temeli ise günümüzün petrolü olarak nitelendirilen bilgidir. 2018 yılında yapılan bir araştırmaya göre tüm dünyadaki bilginin %90'ı sadece son iki yılda üretilmiştir¹. Geçtiğimiz yıl üretilen toplam veri miktarı 79 Zettabyte² olarak tahmin edilirken 2024 yılında bu sayının 147 Zettabyte olacağı tahmin edilmektedir³. SI (International System of Units) tarafından belirlenen ölçüm sistemlerinde Zetta ön ekinden sonra gelen Yotta'nın son olduğu gerçeği ışığında, günümüzde üretilen veri miktarının büyüklüğü daha rahat algılanabilir. Günümüzde mobil cihazlar, sayısal kameralar, uzaktan algılama, yazılım günlükleri (log), RFID cihazlar gibi sayıları giderek artan kaynaklardan veriler toplanmaktadır.

Bu kaynakların sayısının artmasına ek olarak fiziksel nesnelere birbiriyle ve daha büyük sistemlerle haberleşebildiği bir iletişim ağı olarak tanımlanan Nesnelere İnterneti (IoT) ile yakın gelecekte çok daha fazla sayıda küçük ve ucuz cihazdan elde edilecek veriler de düşünülürken önümüzdeki yıllarda veri miktarının katlanarak artması beklenmektedir.

Üretilen veri miktarının katlanarak büyümesi nedeni ile geleneksel veri elde edimi, saklama, iletme ve işleme yöntemleri yetersiz kaldığından bu konularda bir çok yeni çalışma yapılmış ve bu çalışmalar zamanla Büyük Veri (Big Data) tanımını doğurmuştur. Tipik olarak Büyük Veri tanımında Hacim (Volume), Çeşitlilik (Variety), Hız (Velocity), Doğruluk (Veracity) ve Değer Katma (Value) başlıkları Büyük Veriyi tanımlayan 5 temel karakteristik özellik olarak kullanılmaktadır. Bu özelliklerden Hacim, verinin çokluğunu temsil etmektedir. Büyük Verinin büyük olabilmesi için tanımlı bir sınır bulunmamasıyla

* İstanbul Teknik Üniversitesi, Yapay Zeka ve Veri Mühendisliği Bölümü

¹ <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=5c78904f60ba>

² Zettabyte = 10²¹ Byte

³ IDC and Statista Report, 2021, "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025"

beraber genel olarak erişilmesi, saklanması ve yönetilmesi geleneksel yöntemlerle verimli bir şekilde gerçekleştirilemeyen büyüklükleri kastedilmektedir. Çeşitlilik kavramı ile yapısal veri (structured data) ile beraber resim, video, metin gibi yapısal olmayan çeşitli niteliklerdeki verilerin de kullanılması belirtilmektedir. Hızlı üretilen verinin yine hızlı bir şekilde saklanması ve işlenmesi gerekmektedir. Üretilen ve saklanan verinin bütünlüğünün ve doğruluğunun sağlanması Doğruluk olarak tanımlanmaktadır. Tüm bu özelliklerden belki de en önemlisi, yürütülen bütün bu faaliyetlerin işletmeye Değer Katakacak sonuçlar üretebilmesi olarak tanımlanmaktadır.

Çağımızda Büyük Veri ile paralel olarak çığır açan gelişmelerin yaşandığı Veri Bilimi ve Yapay Zekâ alanlarının da Büyük Veri ile yakın ilişki içinde olduğu görülmektedir. Günümüzde ağırlıklı olarak yapay öğrenme (Machine Learning) ve derin öğrenme (Deep Learning) temelli çalışmaların yürütüldüğü yapay zekâ çalışmaları gibi veri bilimi çalışmaları da temelde veriye gereksinim duymaktadır. Büyük miktardaki verinin verimli bir şekilde erişilebilir ve işlenebilir hâle gelmesi ile bu alanlarda da daha gelişmiş ve başarılı sistemler ortaya çıkabilmektedir. Örneğin 40 GB metin verisi⁴ kullanılarak eğitilen bir doğal dil işleme modeli olan GPT-3'te⁵ 175 milyar parametre optimizasyonu yapılmış ve bu model kullanılarak insanlar tarafından üretilen metinlerden ayırt edilmesi çok zor olan metinler otomatik olarak üretilmiştir.

Son yıllarda hemen hemen tüm sektörlerde çarpıcı değişimleri tetikleyen Sayısal Dönüşüm (Digital Transformation) kavramının önemli bir sacayağını oluşturan Büyük Verinin birçok farklı alanda yenilikçi uygulamalarını gözlemlenmektedir. Örneğin imalat sektöründe ekipmanlardan elde edilen verilerin işlenmesi ile bir ekipman (örneğin bir motor) henüz bozulmadan önce yakın zamanda bozulacağına ilişkin uyarıların oluşması ve üretim hattında büyük bir sorun yaşanmadan ilgili ekipmanın değişiminin planlanması hayata geçmiştir. Kestirimsel Bakım (Predictive Maintenance) olarak tanımlanan bu uygulamadan başka üretim optimizasyonu ve işletim verimliliği konularında da Büyük Veri uygulamalarından yararlanılmaktadır. Finans alanında ise sahtekarlıkların tespiti, mevzuat uyumlulukların sağlanması, kara para aklama ve yapay zeka destekli al-sat sistemleri gibi çözümler görülmektedir. Örneğin Two Sigma isimli bir fon⁶ yönetim şirketi, yönetmekte olduğu büyük ölçekteki bir fon için yatırım kararlarını vermek üzere tamamen veri odaklı yöntemler kullanmakta, bu amaçla 10 binden fazla veri kaynağından beslediği ve boyutu her gün artan 71⁷ PB büyüklüğünde bir veri havuzunu kullanmaktadır. Hukuk alanında çalışmalar yürüten bir firma⁸, çok sayıda ticari sözleşmeden oluşan büyük bir veri kümesi kullanarak sözleşmelerdeki hataları tespit eden yapay zeka destekli bir sistemi geliştirmiştir. Yapılan bir çalışmada, 20 deneyimli avukatın ortalama okuma süresi 92 dakika ve başarı oranı %85 iken geliştirilen sistemin okuma süresi 26 saniye ve başarı oranı %94 olarak belirtilmektedir⁹. Sağlık alanında Genom araştırmalarından anlık alarmlar üretilmesine, telekom sektöründe ağ kapasitesinin planlanmasından müşteri kayıplarının belirlenmesine, perakende sektöründe fiyat politikalarının belirlenmesinden mağaza içi müşteri deneyiminin artırılmasına, petrokimya sektöründe yeni sahaların tespit edilmesinden üretimin planlanmasına kadar pek çok farklı sektörde birçok uygulama alanında Büyük Veri temelli çözümlerin gerçekleştirildiği görülmektedir.

Teknolojide yaşanan gelişmelere paralel olarak tarihte bazı meslekler yok olmuş ve yeni çalışma alanları ortaya çıkmıştır. Önceki dönemlere kıyasla çok daha hızlı değişim ve gelişmelerin yaşandığı Bilgi Çağı'nda da benzer durumun tekrar yaşanacağı açık olarak görülmektedir. Bu nedenle gelecek henüz gelmeden, geleceğe dair öngörülerin bu kadar net görüldüğü bir ortamda insan kaynağımızın geleceğin dünyasına hazırlanması gerekmektedir. Bu amaçla Büyük Veri, Yapay Zeka ve Veri Bilimi gibi alanlarda yetişmiş insan kapasitemizin ülke olarak artırılması, bu teknoloji alanlarında akademik ve ticari Ar-Ge faaliyetlerinin artarak desteklenmesi hayati önem arz etmektedir. Ancak belki de bu faaliyetlerden daha da önemlisi, bu alanda yetişmiş insan gücüne tüm dünyanın gereksinim duyması nedeni ile yatırım yapılarak eğitilen bu insan kaynağının yurt dışına göçünü engelleyecek politikaların hayata geçirilmesi olarak görülmektedir.

⁴ 40 GB metin verisi = ~ 8 milyar sözcük (ort. 5 harf/sözcük) = ~ 80.000 kitap (ort. 100.000 sözcük/kitap)

⁵ <https://arxiv.org/abs/2005.14165>

⁶ 2022 yılı itibarıyla yaklaşık 50 Milyar USD

⁷ 71 Petabyte = 71.000 Terabyte = 71.000.000 GB

⁸ LawGeex (www.lawgeex.com)

⁹ <https://blog.lawgeex.com/ai-more-accurate-than-lawyers>

EĞİTİMDE BÜYÜK VERİ KULLANIMI

*Prof. Dr. Hasan KARAL

*Doç. Dr. Ali Kürşat ERÜMİT

İnternetin doğuşuyla birlikte veri üretimi ve aktarımı tarih boyunca görülmemiş bir hızda ve büyüklükte gerçekleşmeye başlamıştır. Bilgisayarların veri depolama kapasitelerinin artması ve bulut teknolojileri ile verinin pek çok noktadan aynı anda üretilebilir ve işlenebilir olması belirli bir konu ile ilgili farklı özelliklere sahip kullanıcılardan veri alabilmeyi ya da belirli kullanıcılar için çok farklı kaynaklardan veri toplayabilmeyi mümkün kılmıştır. Bu durum iletişim araçlarından alışveriş sitelerine, sosyal medyadan sağlık verilerine, konum verisinden bir RFID sinyaline kadar hemen hemen tüm alanlarda kullanıcıların hareketlerinin ve tercihlerinin kayıt altına alınmasını sağlamıştır. Artık bilgisayarların kullanıcı verilerinden hareketle her hangi bir konuda istatistik oluşturması, çıkarımda bulunması ve hatta karar vermesi mümkündür.

Gelişen teknoloji ile hem yazılımsal hem de donanımsal olarak bir kullanıcının tüm internet hareketleri, konum verileri, kamera kayıtları, sosyal medya hareketleri, kredi kartı kullanım durumu, dijital oyun ve sanal dünya kullanım verilerinin yanı sıra hem devlet hem de özel sektör ile paylaştığı tüm sağlık, öğrenim, sosyal güvenlik, adli sicil geçmişi gibi pek çok kişisel verisi de işlenerek kullanıcıların profillerinin oluşturulmasında ve onlara yönelik çıkarımlarda bulunulmasında kullanılabilir. Diğer taraftan çıkarım olarak ifade edilen öneriler odak noktasına alındığında bu önerilere uygun kullanıcı profilleri ve nedenleri de analiz edilebilmektedir.

Ancak günümüzde artık sadece profillere uygun öneri ya da eşleştirme yapmak yeterli gelmemektedir. İnsanlar, oluşturma sürecinde de makinelerden yardım almaktadır. Örneğin yeni çıkarılacak bir ürünün hangi tüketici kitlesine hangi özelliklerde üretilmesi gerektiğine, henüz senaryosu bile olmayan yeni bir film için nerede, hangi zamanda, hangi konuda, hangi oyuncularla ve hatta hangi yönetmenle çekilmesi gerektiğine ya da yeni iş girişimi önerilerine kadar sonuç çıkarımı için pek çok farklı veri kaynağından pek çok bileşenin olduğu oldukça karmaşık süreçleri makineler gerçekleştirebilmektedir. Sadece ilgili alanda uzman olan ve uzun yıllar boyunca konu ile ilgili verileri ve deneyimi bulunan bir kişinin verebileceği kararları günümüzde makineler oldukça hızlı şekilde alabilmektedir.

Peki, insana özgü bir uzmanlık ve deneyimle alınabilecek kararları makineler nasıl ortaya çıkarabiliyor? Elbette burada esas olan unsur veridir. Karar almak için zamanında ve farklı veri kaynaklarından elde edilmiş yeterli miktarda, değerli ve güvenilir veriye ihtiyaç duyulmaktadır. Bu durum diğer bir soruyu beraberinde getirmektedir. Bu veri nasıl oluşturulacak, toplanacak ve derlenerek sonuca dönüşecek?

Günümüzde veri her türlü dijital araçtan elde edilebilmekte ve bulut teknolojileri sayesinde ortak platformlarda depolanabilmektedir. Ancak alışlagelmiş veri tabanı anlayışında bu kadar farklı kaynaktan, farklı formatlarda ve ilişkilendirilmemiş verinin birbiri ile konuşması mümkün değildir. Bu nedenle yakın geçmişe kadar ziyaret edilen siteler, tıklama sayıları, gidilen yerler, konum bilgileri, sağlık ya da öğrenim bilgileri gibi üretilen verilerin oldukça sınırlı şekilde ve sadece kendi bağlamında değerlendirilmekte büyük bir kısmı ise arşiv haline gelmekteydi. Bu durum verinin sadece yığın olarak bağımsız veri tabanlarında saklanmaktan öteye geçememesine neden olmaktadır. Ancak bulut teknolojilerindeki gelişmeler ve veri madenciliği teknikleri verinin oluşumunu, erişilebilirliğini ve birleştirilebilmesini sağlamıştır. Dolayısı ile bu zamana kadar anlamsız görünen veri yığınları artık belirli bir amaç doğrultusunda anlamlandırılabilir.

Belirli bir amaç için her hangi bir anda farklı veri kaynaklarından ve farklı formatlarda toplanabilen yeterli miktarda, değerli ve güvenilir, işlenmiş ya da ham veri kümesine büyük veri adı verilmektedir. Büyük

* Trabzon Üniversitesi Fatih Eğitim Fakültesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü

verinin 3V olarak ta nitelendirilen Hacim (Volume), Çeşitlilik (Variety) ve Hız (Velocity) özelliklerine sahip olması gerekir. Büyük veri kullanımı arttıkça daha kullanılabilir sonuçlar elde edebilmek için bu 3 kavrama Doğruluk (Veracity), Değer (Value) ve Değişkenlik (Variability) özellikleri de eklenmiştir.

Büyük veri, sağlıktan borsaya reklamcılıktan ulaşımına pek çok alanda tahminlerde bulunabilmek için kullanılmaktadır. Bu alanlar içerisinde eğitim, nispeten yeni olmakla birlikte sağlayabileceği avantajlarla büyük bir potansiyele sahiptir.

Büyük veri için daha önce ifade edilen tüm çıkarım imkanlarının farklı mecralardan sağlanabilecek öğrenim verilerinin işlenmesi için kullanımı, başta bireysel öğretim olmak üzere eğitimde geniş bir kullanım alanı sağlamaktadır. Çünkü artık öğrenim sadece okul sınırları içinde değildir. Özellikle Covid-19 pandemisi ile birlikte uzaktan eğitime geçişin hızlanması, dijital öğrenim verilerinin katlanarak artmasını sağlamıştır. Artık öğrencilerin arama geçmişleri, ziyaret ettikleri eğitim platformları, ziyaret sıklıkları, bu platformlarda geçirdikleri süreler, tıklama sayıları, yaptıkları tüm faaliyetler, tıkladıkları içerikler, bu içeriklerdeki ilerleme durumları, içerik tercihleri, bitirdikleri kurslar, çeşitli video izleme platformlarında izledikleri videolar, sosyal medyada eriştikleri içerikler, beğenileri vb. her türlü hareketleri eğitim-öğretim süreçleri için anlamlı bir veriye dönüştürülebilmektedir. Önceden birlikte bir anlam ifade etmeyen bu veriler artık bir çıkarım yapabilmek için anlamlandırılabilir. Böylece bu veriler genel olarak öğrencilere eğitim içeriği ve çeşidi önermek, ders/bölüm/okul önerileri yapmak, mesleki anlamda yönlendirmek, öğrenme eksiklerinin belirlenmesi ve bu eksiklerin giderilebilmesine yönelik yönlendirmeler yapmak gibi amaçlarla kullanılmaktadır.

Öğrencilerin akademik başarılarının belirlenmesi için büyük veri kullanımı, şimdiye kadarki tek düze sınav, ödev, proje, performans değerlendirmesinin ötesine geçerek derse katılım ve ders için harcanan çabanın da farklı kaynaklardan elde edilecek verilerle ölçülmesine olanak sağlayabilmektedir. Örneğin belirli bir dersi alan öğrencilerin belirli platformlarda yaptıkları yorumlar, beğeni sayıları, arama seçenekleri, geçirdikleri süre vb. internet verileri incelenerek derse yönelik ilgilerinin ölçülmesi ya da bir öğretmenin performansının belirlenmesi sağlanabilmektedir.

Eğitimde büyük veri, derslerin daha etkin hale getirilebilmesi amacıyla da kullanılmaktadır. Örneğin öğrencilerin sınavlarda başarısız olduğu konular, hata sayıları, çözüm adımı sayıları ve harcadıkları süre verileri eğitimcilerin ders için paylaştıkları içeriklerin yeterliklerinin belirlenmesinde kullanılabilir. Öğrencilerden alınacak ayrıntılı girdiler ile içeriğin yeterliliğinin belirlenmesi ve eksiklerinin giderilmesi ile en uygun içeriklerin oluşturulması sağlanabilmektedir. Bu yolla öğrenci ve eğitimci verileri öğretim tasarımının iyileştirilmesi amacıyla da kullanılabilir.

Eğitim-öğretim ile ilgili süreçlerini yalnızca dersler boyutunda düşünmemek gerekmektedir. Örneğin öğrencilerin web geçmişlerinden hareketle ilgi alanlarının belirlenerek onlara yönelik sosyal, sportif ya da sanatsal etkinlikler düzenlenebilir. Aynı durum bireysel olarak öğrenciler için ilgi alanlarına uygun faaliyetlere yönlendirmelerinde kullanılabilir.

Çeşitli alışveriş ve video akış sitelerinin müşterilerine öneri amaçlı yaptığı çıkarımlar öğrencilerin web hareketlerinin izlenmesi yoluyla uygun eğitim içeriklerinin getirilmesinde kullanılmaktadır. Web hareketi verisine, öğrencinin akademik başarısını, daha önce ziyaret ettiği yerleri, önceki başarılarını, deneyimini ve ilgi alanlarını da ekleyerek üniversite ve bölüm tercihi ya da sonraki aşamada meslek tercihi için çıkarımlarda bulunabilmektedir.

Eğitimde büyük veri kullanımı sadece öğrenciler değil kurumlar için de yaygınlaşmaktadır. Örneğin eğitim kurumları daha gerçekçi ve tarafsız sonuçlar elde edebilmek amacıyla öğrenci ve personel memnuniyetlerini anketler yerine sosyal medya ve web platformlarından gelen yorumları analiz ederek belirleyebilmektedir. Yine kurumlar verimlilik içinde büyük veriyi kullanabilmektedir. Örneğin kampüsteki ağ günlüklerini izleyerek ağ trafiğini düzenleme ya da öğrenci tercihlerini analiz ederek verilen hizmetle ilgili tüm bileşenlerin yeterliklerinin belirlenmesi sağlanabilir. Yine kampüs içi verilerden öğrencilerin ihtiyaçları belirlenebilir.

Eğitimde büyük veri çalışmalarının görülmeye başlandığı 2006 yılından itibaren çalışma sayısında katlanarak artan bir ivme görülmektedir. Özellikle 2014 yılından günümüze yoğunlaşan büyük veri çalışmalarının; eğitim-öğretim sürecinde yer alan aktörlerin üretebildiği veri türlerin çeşitlenmesi ve

veri üretme ve paylaşma hızlarının artmasıyla daha kapsamlı ve yaygın olacağı aşikârdır. Bu çalışmalardan elde edilecek verilerin ise özellikle bu süreçlerdeki eksikleri, ihtiyaçları ve yapılması gerekenleri belirlemek açısından önemi daha iyi anlaşılacaktır. Artık sadece insana özgü tecrübeyle elde edilen verileri kullanarak karar verme süreci yerini geçmişten günümüze gelen ve hatta o anda oluşan çok farklı kaynaklardan verileri değerlendirerek karar verebilen sistemlere bırakmaya başlamıştır. Tüm alanlarda kendini gösteren bu durum eğitim öğretim süreçleri içinde kaçınılmaz olarak önümüzdeki yıllarda büyük veri kavramından daha çok bahsedilmesine neden olacaktır.

BÜYÜK VERİ ÖĞRETİMİNDE İLK ADIM: GEREKSİNİMLER VE MİMARİ KATMANLAR

*Doç. Dr. Adnan Özsoy

Büyük veri mimarisi, geleneksel veri tabanı sistemleri için çok karmaşık veya büyük olan verilerin işlenmesini, alınmasını ve analizini gerçekleştirebilir. İş amaçları için analiz edilecek büyük miktarda veriyi yönetmek, veri analitiğini yönlendirmek ve büyük veri analitiği araçlarının hayati iş bilgilerini çıkarabileceği bir ortam sağlamak için kullanılan kapsayıcı sistemdir, ayrıca büyük veri mimarisi çerçevesi bir referans planı olarak hizmet eder. Büyük Veri araçlarını öğrenmede genel jargon olarak gereksinimler ve ana taslak mimarinin bilinmesi ilk aşama olarak değerlendirilebilir.

Büyük Veri Gereksinimleri

Büyük Veri için öncelikle gereksinimlerin bilinmesi gerekir. Bu gereksinimleri veri gereksinimleri ve işleme gereksinimleri olarak ikiye bölebiliriz. Veri gereksinimlerine bakıldığında üç ana başlık altında veri gereksinimlerini inceleyebiliriz:

1-Verinin Yapısı: Bildiğiniz gibi, veriler tablolarda veya dosyalarda saklanabilir. Veriler önceden tanımlanmış bir veri modelinde depolanıyorsa (yani bir şeması varsa), buna yapılandırılmış veri denir. Dosyalarda saklanıyorsa ve önceden tanımlanmış bir modeli yoksa, yapılandırılmamış veri olarak adlandırılır.

Bu bağlamda iki tür yapıdan söz edebiliriz: Yapılandırılmış/Yapılandırılmamış. Bazı kaynaklarda bu ikisinin birleşimi olarak Yarı Yapılandırılmış ifadesi de geçmektedir.

2-Boyut: Boyut ile veri miktarını değerlendiririz.

3-Kullanıcı ve Kaynak (sink and source) Çıktısı: Hangi oranda verinin sisteme kabul edilebileceğini, hangi hızda güncellenebileceğini ve sisteme dönüştürülebileceğini tanımlar.

Diğer bir gereksinim ise işleme gereksinimleridir. Bu gereksinimleri de üç ana başlık altında inceleyebiliriz.

1-İşlem süresi: Verileri işlemek için gereken süre

2-Sorgu zamanı: Bir sistemin sorguları yürütmesi için geçen süre.

Bu ilk iki başlık t-shirt boyutları dediğimiz S-M-L-XL gibi ölçütlerle büyüklükleri ifade edilir.

3-Kesinlik: Veri işlemenin doğruluğu açısından iki ana şekilde, tam bir doğruluk ya da yaklaşık değer üzerinden bir kesinlik sunulmasıdır.

Büyük Veri Mimarisi Katmanları

Gereksinimlere bağlı olarak Büyük Veri Mimarisi geliştirme ve uygulamalardaki ana katmanları tanımlar. Farklı kaynaklarda dört ile sekiz arasında farklı sayıda katmanlar ifade edilmiş olsa da bu yazıda altı ana katmandan bahsedilecektir.

* Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü

Toplama ve İşleme Katmanı

Bu katman farklı kaynaklardan gelen veriler için ilk adımdır. Veri alımı, verilerin önceliklendirildiği ve kategorilere ayrıldığı anlamına gelir ve bu alım katmanından veri hattının geri kalanına taşınmasına odaklanılır. Analitik yeteneklerin başlayabilmesi için bileşenlerin ayrıştırıldığı veri mimarisi katmanıdır. Bu birincil katmanda odak, topladığımız verilerin analitik yeteneklerin başlayabilmesi için bileşenlerin ayrıştırılması ve veri hattı işleme sisteminde uzmanlaşılmasıdır. Bu kapsamda kullanılan en yaygın araçlar Apache Sqoop, Apache Storm, Apache Spark, Hadoop ve Apache Flink araçlarıdır.

Veri Depolama Katmanı

Bir sonraki aşamada veri alma süreci akışının ana sorunu, kullanıma göre verileri doğru yerde saklamaktır. Veriyi uzun yıllar boyunca depolamak için ilişkisel veri tabanları kullanılmıştır ancak çeşitlenen ve artan veri kaynakları ilişkisel veri olma zorunluluğunu ortadan kaldırmıştır. Farklı veri çeşitlerini işlemek için farklı veri tabanlarına ihtiyacımız vardır, ancak farklı veri tabanları kullanmak ek yük yaratmaktadır. Büyük Veriyi depolamak için kullanılan farklı veri depolama araçları türleri vardır. HDFS (Hadoop Dağıtılmış Dosya Sistemi), GlusterFS (Güvenilir Dağıtılmış Dosya Sistemi), ve Amazon S3 Depolama Hizmeti gibi hizmetler sayılabilir.

Veri Sorgu Katmanı

Aktif analitik işleminin gerçekleştiği veri mimarisi katmanıdır. Bu, etkileşimli sorguların gerekli olduğu bir alandır ve geleneksel olarak SQL uzman geliştiricilerinin egemen olduğu bir bölgedir. Hadoop gibi araçlar uzun bir analiz süreci gerektirdiğinden depolama alanı yetersiz kalabilmektedir. Bu nedenle veri, ayıklama-dönüştürme-yükleme (Extract-Transform-Load) adımlarından geçer ve ardından bir veri tabanına veya veri ambarına koyulur. Veri alımı ve veri analitiği, bir veri alımı çerçevesini oluştururken aynı zamanda bu kadar büyük miktarda verinin hesaplanırken oluşan sorunları çözen bir adım haline gelmiştir. Bu kapsamda kullanılan en yaygın araçlar Apache Hive, Apache Spark SQL, Amazon Redshift, Presto – Büyük Veri İçin SQL Sorgu Motoru araçlarıdır.

Veri Görselleştirme Katmanı

Görselleştirme katmanı, projenin başarısını ölçen bir göstergedir. Bu, kullanıcının algıladığı verinin değeridir. Hacimli miktarda veriyi işlemeye ve depolamaya yardımcı olmakla birlikte, Hadoop ve diğer araçlar, veri görselleştirme ve bilgi dağıtımı için yerleşik bir bilgiye sahip değildir ve veri alımı işleme hattındaki son iş kullanıcıları tarafından kolayca anlaşılabilir hale getirme adına görselleştirme kullanılır. Veri Görselleştirme adına pano (dashboard) yapıları gerçek zamanlı görselleştirme çokça tercih edilmektedir. Tableau aracı ile veri görselleştirme sürükle ve bırak işleviyle piyasada bulunan en zengin veri görselleştirme aracıdır. Kibana da veri kümelerini keşfetmekte kullanılan araçlardan biridir. Bunların yanı sıra akıllı ajanlar, öneri sistemleri, kodlama açısından Angular.JS Javascript çerçevesi ile Tek Sayfa Uygulaması (SPA) projelerinde kullanılan uygulama geliştirme araçlarıdır.

Veri Güvenliği ve Veri Akış Katmanı

Güvenlik, her tür verinin çok önemli bir parçasıdır ve ayrıca Büyük Veri Mimarisi'nin önemli bir yönüdür. Veri alımından başlayarak depolama, analiz, keşif ve tüketime kadar verinin tüm katmanlarında güvenlik uygulanması gerekir. Veri hattına veri alımında güvenlik sağlamak için birkaç adım vardır. Bu adımlar kimlik doğrulaması, giriş kontrolü, şifreleme ve veri maskeleyme, veri erişiminin denetlenmesi olarak sayılabilir.

Veri İzleme Katmanı

Kurumsal sistemlerdeki verilerin sürekli olarak yenilenmesi, özellikle stratejik ve operasyonel kararların üzerinden alındığı verilerde önemli bir gerekliliktir. Veri alma süreci akışında bol miktarda veri olabilir, ancak değerli olması için güvenilir ve tüketilebilir olması gerekir. Kuruluşların odak noktası genellikle

büyük miktarda veriyi depolamak ve analiz etmek olsa da, bu verileri olabildiğince güncel tutmak da önemlidir. Bu kapsamda verileri izlemek, denetlemek, test etmek, yönetmek ve kontrol etmek adımları uygulanır. Verilerin sürekli izlenmesi, yönetim mekanizmalarının önemli bir parçasıdır. Bu kapsamda kullanılan araçlardan birkaçı ise Apache Flume, Kafka, Talend, Logstash'dir.

Sonuç

Büyük Veri araçlarını öğreniminde birçok farklı senaryoda birçok farklı araç kullanılmaktadır. Özellikle ilgili katmana bağlı olarak gereksinimler ve yapılacak faaliyetler farklılık gösterir. Bu kapsamda Büyük Veri mimarisinin iyi anlaşılması ve değerlendirilmesi, katmanlar arası geçişler ve birbirlerine girdi-çıkı bağlantısı bulunan uygulama yığınının oluşturulması açısından önem arz etmektedir. Bu katmanlar temelinde oluşturulacak bir ders içeriği ile öğrencilere bu katmanların işlevlerinin anlatılması ve devamında katmanlar arası geçişler için de ilgili uygulamalardan bir uygulama yığı ile örnekler sunulması yararlı olacaktır.

VERİMLİ ÖĞRENME İÇİN BÜYÜK VERİ KULLANAN ARAMA VE ÖNERİ SİSTEMLERİ

*Doç. Dr. İsmail Sengör Altıngövde

Büyük veri, son yıllarda günlük ve profesyonel hayatın dijitalleşmesinin doğal bir sonucu olarak hayatımıza giren bir kavram. Veriyi tanımlayan sadece hacimsel büyüklüğü değil, yüksek akış hızı ve çeşitliliği başta olmak üzere başka özellikleri de var; ama büyük veri denildiğinde akla gelen sadece verinin kendisinden de ibaret değil. Büyük veri kavramı, bu verinin elde edilmesini, saklanmasını ve en nihayetinde işlenerek yeni bilgiye—ve özellikle de karar verme mekanizmalarını yönlendirebilecek bir alt-zekaya—dönüştürülmesini sağlayacak tüm yöntem ve araçları da kapsayan daha büyük bir şemsiye olarak düşünülmeli.

Büyük veri, son dönemde giderek önemi ve etkisi artan e-öğrenme ve teknoloji ile zenginleştirilmiş öğrenme gibi alanlarla da yakından ilişkili; ve örgün eğitim-öğretim kurumlarından kitlesel çevrim-içi ders sitelerine kadar her türlü platform ve her seviyede öğrenci için öğrenme verimliliğini artıracak önemli faydalar sağlayabilir. Bu yazıda odak noktamız ise, üniversite eğitimi seviyesinde öğrenme sürecinde verimliliğin artırılması için büyük veriden faydalanan bilgi erişim yöntemleri, özellikle de arama ve öneri sistemleri. Bilgiye erişim konusu önemli, çünkü artık dünyada ve ülkemizde pek çok üniversitede derslerin yürütülmesi sırasında çok fazla sayıda ders videosu, sunu dosyası, etkileşimli örnek, ders notu, animasyon, belge ve benzeri kaynak—kısaca öğrenme nesnesi—üretiliyor ve ağ üzerinden paylaşılıyor, yani işin doğası gereği yükseköğretim süreci/kurumları büyük veri doğuruyor. Bunun yanı sıra üniversitelerde öğrencilerin kişisel bilgileri, ilgi alanları, geçmişte aldıkları dersler ve bu derslerden aldıkları notlar gibi daha iyi yapılandırılmış ama işlenmeyi ve anlamlandırılmayı bekleyen bir başka büyük veri deposu daha var. Buradaki araştırma sorusu ise öğrenme nesnelere ilişkin bu okyanusta, bir kullanıcının ilgi alanlarını, bilgi birikimini, içinde bulunduğu ortamı/koşulları dikkate alarak onun için öğrenme sürecinde en faydalı olacak öğrenme nesnelere ulaşmak veya önerilebilir.

Arama ve öneri sistemleri bilgisayar bilimi literatüründe sıklıkla çalışılan ve adeta aynı madalyonun iki yüzü olan gibi olan araştırma konularıdır. Arama sistemlerinde kullanıcı genellikle kelime-tabanlı bir sorgu yapar (örneğin, “doğrusal cebir”), ve sistemin hedefi eldeki öğrenme nesnelere arasında bu sorguya en ilgili olanları bulabilmektir. Buradaki temel yaklaşım, sorgu metni ile öğrenme nesnesini oluşturan (veya tanımlayan) metin arasındaki yakınlığı ölçen bir fonksiyon kullanarak sorguya en yakın, diyelim ki ilk-10, nesnenin belirlenmesi ve kullanıcıya gösterilmesidir.

Elbette bu eşleştirme fonksiyonu ham öğrenme nesnelere yerine bunların üzerinde oluşturulmuş bir ters indeks yapısı üzerinde çalıştırılır, böylece sistemde bulunan tüm nesnelere yerine sadece sorgu kelimelerinden en az bir tanesini içeren nesnelere için yakınlık skoru hesaplanır.

Yukarıda özetlenen basit yaklaşım, günümüzdeki modern arama sistemleri için sorgu işlememenin ilk aşamasıdır ve sadece daha küçük sayıda, mesela bir kaç bin tane, aday nesne bulmak için kullanılabilir. İkinci aşamada, bu nesnelere metinsel yakınlığın ötesine geçerek daha rafine skorlar atanması gerekir ki, burası büyük veri ve makine öğrenmesinin devreye girdiği yerdir. Bu aşamada, eldeki büyük veriden, yani öğrenme nesnelere, kullanıcılar ve sorgulardan (ve bunların birbirleriyle ilişkilerinden) elde edilen çok sayıda öznitelik ile eğitilmiş makine öğrenmesi modelleri kullanılır. Bu modellerde kullanılan öznitelikler sayesinde bir öğrenme nesnesine verilen skor, hem bütüncül bir bakış açısından eldeki nesnenin

* Orta Doğu Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü

sorguyla ilişkisini (örneğin, sorgunun ilgili olduğu derse/konuya kavramsal yakınlığı, bu konudaki geçmişteki sorgular için başka kullanıcılar tarafından tercih edilme sıklığı, vb.) hem de o kullanıcıya özel (kişiselleştirilmiş) faydasını (örneğin, kullanıcının sistemdeki geçmiş davranışlarına göre ilgilendiği konulara yakınlığı, kullanıcının tercih ettiği zorluk seviyesinde ve veri türünde olması, kullanıcının bu konudaki bilgi ve başarı seviyesine uygunluğu, vb.) temsil edebilir. Tabii ki bu özniteliklerin sadece metin verisinden çıkartılmış olması gerekmez, video, görsel veya başka veri tipindeki öğrenme nesnelere de elde edilmiş olabilirler. Günümüzde derin sinir ağları kullanılarak elde edilen çeşitli kestirimler de bahsedilen modellerde öznitelik olarak kullanılabilir. Örneğin, sorgu ve öğrenme nesnesi arasındaki anlamsal yakınlık büyük veri üzerinde eğitilmiş dil modelleriyle (BERT gibi) çok daha iyi ölçülebilir. Ayrıca bu çeşit modellerin, temel arama işlevinin ötesine geçip, sorulan bir soruya direkt cevap bulmak gibi daha iddialı bir görev için de oldukça başarılı olduğu yakın zamanlı bilimsel çalışmalarda gösterilmiştir.

Madalyonun diğer yüzü olan öneri sistemlerinde ise girdi olan bir sorgu yoktur, bir bakıma kullanıcının kendisi sorgudur; amaç ise eldeki kullanıcı profiline en uygun öğrenme nesnelere tavsiye etmektir. Örneğin, bir kullanıcı sistem üzerinde "doğrusal cebir" konusunda giriş seviyesinde öğrenme nesnelere erişmişse, bu konudaki bilgisini ölçen veya bir sonraki seviyedeki konuyu anlatan nesnelere önerilebilir, hatta daha akıllı bir sistem kullanıcının öncelikli ilgi alanını algılayarak (örneğin, bilgisayar bilimi) özellikle bu alandaki doğrusal cebir uygulamalarına dair öğrenme nesnelere önerebilir. Modern öneri sistemleri de yine büyük ölçekli veriler üzerinde makine öğrenmesi yöntemleriyle elde edilmiş modellerden faydalanmaktadır. İçerik tabanlı öneri sistemlerinde, kullanıcının geçmişte eriştiği öğrenme nesnelere en yakın diğer nesnelere önerilir; buradaki yakınlık fonksiyonu nesnelere içeriklerinden elde edilmiş pek çok öznitelik üzerinde tanımlanmış olabilir ve bu özniteliklerin bir kısmı da yine daha başka veri madenciliği (kümeleme gibi) veya makine öğrenmesi (sınıflandırma gibi) yöntemleriyle elde edilmiş olabilir. İşbirlikçi filtreleme yapan öneri sistemlerinde ise kullanıcılar ve geçmişte eriştikleri öğrenme nesnelere bilgisinden yararlanır, ve eldeki kullanıcıya en çok benzeyen diğer kullanıcıların eriştiği nesnelere önerilebilir. Bu amaçla kullanıcı-nesne matrisi üzerinde çarpanlara ayırma yöntemi uygulanarak her kullanıcı ve nesne daha küçük boyutlu bir vektörle temsil edilir; bu gösterimde aynı konuyla ilgili nesnelere belirli bir vektör elemanı yüksek, diğer elemanları ise görece düşük değerler alabilmektedir (yani, bir bakıma vektör elemanları nesnelere farklı konulara veya kullanıcıların farklı ilgi alanlarına karşılık gelmektedir). Günümüzde bu yöntemin ana fikrinden hareketle kullanıcı veya nesnelere gömme vektörleriyle temsili tercih edilebilmektedir; burada yine büyük ölçekli veri ve derin sinir ağları kullanılarak kullanıcıların ve nesnelere çok çeşitli özniteliklerinden gömme vektörleri elde edilmekte, sonra bu vektörler arası yakınlık hesaplanarak eldeki kullanıcıya önerilecek nesnelere bulunmaktadır.

Sonuç olarak diyebiliriz ki, yükseköğretim alanında bir taraftan çok büyük hacimde öğrenme nesnesi hızla ve çeşitli veri türlerinde üretilirken, diğer taraftan da bu nesnelere erişen kullanıcılar hakkında yüksek miktarda ve zengin içerikli veri biriktirmektedir. Tüm bu veriler üzerinde geliştirilecek akıllı arama ve öneri sistemleriyle, her kullanıcı için en faydalı olacak (kişiselleştirilmiş) şekilde öğrenme nesnelere erişimi sağlamak ve böylece öğrenme verimini artırmak artık ulaşılabilir bir hedef olarak ortaya çıkmaktadır. Kim bilir, öğrenmek istediğimiz bir konunun sadece başlığını yazdığımızda, sistemin bilgi birikimine en uygun öğrenme nesnelere, yine en rahat özümseyeceğimiz sırada ve veri türünde sunacağı "otomatik-derlenmiş özel dersler" belki de artık bir adım ötemizdedir...

BÜYÜK VERİ VE VERİ BİLİMİNİN YÜKSEKÖĞRETİMDEKİ OLASI KULLANIM ALANLARI

*Dr. Öğretim Üyesi Hamdi Dibeklioğlu

İçinde yaşadığımız bilişim çağında, verinin üretilme hızı çok yüksek ve bu hız katlanarak artıyor. Cep telefonumuzla çektiğimiz fotoğraflardan, e-ticaret alverişlerimizin faturalarına, e-devlet uygulamaları üzerinden yaptığımız başvurulardan, hastane kayıtlarımıza kadar herşey aslında birer sayısal veri. İşte tüm bu verilerin sürekli şekilde üretilip depolanması son yıllarda sıklıkla duyduğumuz büyük veri kavramını ortaya çıkarıyor. Verinin, bu denli yüksek miktarda üretildiği bir dönemde, etkin şekilde incele-nip yorumlanarak bilgi çıkarımında kullanılabilmesi büyük önem arz ediyor. Aksi halde, hiç durmadan üretilen verinin sadece kayıt amaçlı depolanması onu nitelsiz hale getiriyor. Bu sebeple, istatistik, yap-ay öğrenme ve veri analizi gibi bir çok alanı bir araya getirerek verinin anlamlandırılmasına odaklanan veri bilimi de her geçen gün önemini arttırıyor. Peki, yükseköğretimde büyük veri ve veri biliminden nasıl yararlanabiliriz?

Üniversitelerde, öğrencilerle ilgili birçok bilgi düzenli şekilde kaydedilmekte. Örneğin, alınan dersler, bu derslerin hangi dönemde alındığı ve notları, her ders için devamsızlık kayıtları, mezun olma ya da olam-ama durumu, mezun olunduysa mezuniyet tarihi ve ortalaması, mezun olunamadıysa ayrılma tarihi gibi kayıtların yıllar boyunca tüm öğrenciler için tutulduğu düşünülürken üniversitelerde ne denli büyük veri üretildiği anlaşılabilir. Bu verilerin sadece arşivleme amacıyla tutulması net bir yarar sağlamamanın yanı sıra depolama alanı gereksinimleri sebebiyle bir külfet olarak da görülebilir. Öte yandan, bu tip veriler gelecek nesillerin en iyi şekilde yetiştirilmesinde kullanılabilir bir hazineye de dönüştürülebil-ir. Az önce bahsettiğimiz öğrenci kayıtlarını inceleyerek, başarılı ve başarısız öğrencilerin karakteristik özelliklerini ortaya çıkarmak mümkün olabilir. Bu karakteristikleri yapay öğrenme algoritmaları ile modelleyerek öğrenimine devam eden öğrencilerin gelecekteki olası başarısızlıklarını önceden tahmin edebilir ve gerekli desteği sağlayabiliriz; nasıl mı? Diyelim ki "A" dersinde yüksek ihtimalle başarısız olacak öğrencileri tespit etmek istiyoruz. Geçmiş yıllarda bu derste başarılı ve başarısız olmuş öğrencilerin kayıtlarını inceleyerek, "A" dersinden önce hangi dersleri hangi sırayla almış olduklarını, bu derslerdeki not ve devamsızlıklarını istatistiksel olarak modelleyebiliriz. Bu şekilde, eski veriler kullanılarak, "A" dersinde başarısız olmuş öğrencilerin başarılı olanlardan farkları yani tanımlayıcı özellikleri öğrenilebilir ve benzer özellikleri gösteren öğrenciler daha bu dersi almadan uyarılabilir. Dahası, "A" dersinden başarılı olmuş öğrencilerin ortak özellikleri incelenerek, bu derste başarısız olabileceği tahminlenen öğrencilere nelere dikkat etmeleri konusunda detaylı bilgi verilebilir.

Farklı bir örnek olarak, başladığı lisans programını tamamlayamayan ve üniversiteden ayrılan öğrencilerin verileri çözümlenerek, benzer örüntüleri sergileyen öğrencilere çok geç kalınmadan yapıcı tavsiyelerde bulunulabilir.

Veri ne kadar çeşitli ve detaylı ise veriden yararlı bilgi çıkarılabilme şansımız da o kadar artar. Mese-la, lisans eğitimi sonrasında mezunların hangi sektörlerde ve hangi pozisyonlarda çalıştığı, mezunlar öğrenimlerine yüksek lisans ve doktora ile devam ettilerse hangi öğretim programlarına katıldıkları gibi bilgilerin de ilgili üniversiteler tarafından tutulduğunu düşünelim. Böyle bir senaryoda öğretim başarısını iyileştirmenin yanı sıra, öğrencilerin ulaşmak istedikleri kariyere de doğru şekilde yönlendirilebileceği öngörülebilir. Önceki örneklerimizdeki gibi, benzer kariyerleri tercih eden mezunların öğrenimleri sırasın-daki ortak özellikleri çözümlenip ortaya çıkarılarak öğrencilerin hedefledikleri kariyere ulaşabilmesi konusunda yardımcı olunabilir. Öte yandan buradaki temel sorun, mezuniyet sonrası kariyer ve devam

* Bilkent Üniversitesi, Bilgisayar Mühendisliği Bölümü

eden yükseköğrenim bilgilerinin de elde edilerek öğrenci verileri ile ilişkilendirilebilmesidir. Bir başka deęişle, her kişinin öğrenim ve kariyer verileri bağlantılı şekilde tutulmalıdır.

Özetlemek gerekirse, öğrenci verilerinin analizi ve modellenmesi ile yükseköğretimde öğrenci performansını arttıracak ve kariyer planlamasında yardımcı olabilecek sistemlerin geliştirilebilmesi mümkündür. Geliştirilecek model ve sistemlerin doğruluk ve başarımının, veri miktarı ve çeşitlilięi ile doğru orantılı şekilde artacağı öngörülebilir. Bu bağlamda, üniversitelerde öğrenci verilerinin detaylı ve düzenli bir şekilde tutulması büyük önem arz etmektedir. Uzun vadede, üniversiteler ve dięer kurumlar arası veri paylaşımına ve bu verilerin birlikte kullanımına olanak vermek adına, kaydı tutulacak bilgilerin ve kullanılacak veri formatlarının ilgili kurumlar tarafından ortak şekilde belirlenmesi yararlı olacaktır.





Yükseköğretim Kurulu Başkanlığı 06539 Bilkent / Ankara-Türkiye
Telefon: +90 (312) 298 70 00 **Faks:** +90 (312) 266 47 59
www.yok.gov.tr