DOSYA: BÜYÜK VERİ

Dosya Editörü: Hasan DAĞ

Khas öğretim üyesi

Son yıllarda üretilen verinin neredeyse insanlığın yazılı tarihine eşit miktara ulaşmasıyla ortaya çıkan ‘‘Büyük Veri’’ kavramı, gelecekte bizi bekleyen dönüştürücü teknolojiler, bunların veriye etkileri ile siyasetin veriyle değişen çehresi masaya yatırılıyor.



BÜYÜK VERİ DOSYASINI AÇARKEN

Hasan DAĞ

Khas öğretim üyesi



Günümüzde yüzlerce milyon insan fotoğraf çekiyor, video kaydediyor, mesaj yazıyor, elektronik posta gönderiyor ve sesli ve/veya görüntülü görüşme gerçekleştiriyor. Dünya çapında işletmeler müşterilerin tercihleri, yaptıkları alış-veriş ve eğilimleri hakkında veri topluyor. Belediyeler araç takip sistemleri ve sensörler aracılığı ile ulaşım başta olmak üzere tüm hizmetleri hakkında veri topluyor. Hükümetler düzenli olarak seçimler, nüfus sayımları, ticaret (ithalat, ihracat), eğitim, sağlık, kaza ve cinayetler gibi akla gelebilecek her alanda veri topluyor. 2013'de dünyadaki toplam veri miktarı 4.4 zettabyte (yani, 4.4 trilyon gigabytes) idi. Bunun 2020'de 44 zettabyte olması öngörülüyor, zira 2020'ye geldiğimizde nesnelerin interneti aracılığı ile birbirine bağlı cihaz sayısının 50 milyar olması beklenmekte ve sensörlerin özellikle Endüstri 4.0 amaçlı kullanımı ile veri oluşumunun katlanarak artacağı tahmin ediliyor.

Verinin büyüklüğünü daha net anlatabilmek için, insanlığın yazıya geçişinden 2000'li yılların başına kadar oluşturduğu verilerin toplamına denk bir miktarda verinin yalnızca 2008 yılında üretildiğini söylemek yeterli olacaktır.

Peki bu kadar büyük veri ne işe yarar ki? Veri çöplükleri mi oluşacak, yoksa veriden bilgi üreten ve bu bilgiyi her türlü ürün ve hizmet geliştirmekte kullanan firmalar/ülkeler rekabette en ön sıralarda mı yer alacak? Bunun en güzel yanıtı bir Alman yazar tarafından aşağıdaki şekilde öngörülmüş bile. Udo Gollub, geçtiğimiz günlerde korkutucu ve bir o kadar da hayatımızı etkileyebilecek, geleceğe dair öngörülerini paylaştığı bir yazı yayınladı. Mercedes Benz'in müdürü, yakın zamanda verdiği bir röportajda; artık rakiplerinin diğer araba firmaları değil Tesla, Amazon, Google, Apple ve benzeri teknoloji şirketleri olduğunu söyledi. Asla değişmeyen üç şeyin olduğu söylenir: ölüm, vergiler ve değişim! Önümüzdeki 5-10 yılda yazılımlar, geleneksel endüstrilerin çoğunluğunu rahatsız edecek. Uber, sadece bir uygulama ve bir tane bile arabası yok ancak yine de dünyanın en büyük taksi şirketi konumunda. Airbnb, dünyanın en büyük otel şirketi ancak hiçbir mülkü yok.

Yapay Zekâ Gittikçe Daha Başarılı!

Bilgisayarlar dünyamızı anlamada gittikçe daha da iyi oluyorlar. 2016'da bir bilgisayar dünyanın en iyi "Go" oyuncusunu yendi ki bu beklenenden 10 yıl daha önce gerçekleşti.

ABD'deki genç avukatların iş bulması çok zor hale geldi. Çünkü IBM'in Watson uygulaması sayesinde sadece saniyeler içinde yasal danışmanlık alabiliyorsunuz ve bunu %90 oranında başarıyla yapıyor ki insan eliyle yapılanların başarı oranı %70. Eğer hukuk okuyorsanız, en kısa zamanda bırakın çünkü gelecekte %90 daha az avukata ihtiyaç olacak ve sadece uzmanlar kalacak piyasada. Watson, insan hemşirelere kıyasla 4 kat daha başarılı bir şekilde kanser teşhisinde bulunabiliyor. Facebook'un sahip olduğu yüz tanıma uygulaması, yüzleri insanlardan daha başarılı bir şekilde tanıyabiliyor. 2030'da, bilgisayarlar insanlardan daha akıllı olacaklar.

Otonom (Özerk) Arabalar

2018'de, kendi kendine giden arabalar ilk kez halka sunulacak. 2020 civarında tüm endüstri bundan kötü etkilenecek. Artık bir araba sahibi olmak istemeyeceksiniz. Telefonunuzla bir araba çağıracaksınız ve olduğunuz yere gelip sizi istediğiniz yere götürecek. Arabayı park etmenize gerek olmayacak, sadece ücreti ödeyeceksiniz ve yolculuk esnasında çalışabileceksiniz. Çocuklarımız asla ehliyet ve araba sahibi olmayacaklar.

Bu durum şehirleri de değiştirecek çünkü %90-95 daha az arabaya ihtiyaç duyacağız. Otoparkları, park alanlarına çevirebiliriz artık. Tüm dünyada her yıl yaklaşık 1,2 milyon insan araba kazalarında hayatını kaybediyor. Günümüzde 100 kilometrede 1 kaza anlamına geliyor bu. Özerk arabalar bu oranı 10 milyon kilometrede 1 kazaya düşürecek. 1 milyondan fazla insanın hayatı kurtulmuş olacak her yıl.

Araba firmalarının çoğunluğu da büyük ihtimalle batmış olacak. Geleneksel araba firmaları evrimsel bir yolla daha iyi arabalar üretmeye çalışıyor ancak teknoloji firmaları (Tesla, Apple, Google…) devrimsel bir yolu tercih edecekler ve tekerlekleri olan bir bilgisayar inşa edecekler.

Bu öngörülerin bir çoğunun gerçekleşeceğine dair hiçbir kuşkum yok, ancak zamanlaması değişebilir. Bu öngörünün neden gerçekleşebileceğine dair iyi bir örnek Whatsapp yazılımıdır. 52 çalışana ve 10 yıldan daha az bir geçmişe sahip olan firma Whatsapp yazılımını 18,5 milyar ABD dolarına satmayı başardı. Peki, bugün bir çoğumuzun vazgeçilmezi olan Whatsapp'ı satın alan firmanın bu yazılımı ücretsiz olarak kullandırmaya devam etmesinin nedeni nedir? Neden çok basit, veri toplamak. Kim kiminle nerede, neyi, ne sıklıkla, niçin paylaşıyor? Toplanan verilerden üretilen bilgiler daha sonra özellikle reklam verecek firmalara satılmakta hiçbir zorlukla karşılaşmıyor, bilakis deyim yerindeyse kapış kapış gidiyor.

Genelde "veri"nin, özelde ise "büyük veri"nin öneminin dünya çapında artması nedeniyle hazırlanan "Büyük Veri" dosyası konunun farklı boyutlarını ele alan dört yazı ile karşınızda. Keyifli okumalar...

BÜYÜK VERİ YOLCULUĞU

Hasan DAĞ

Khas öğretim üyesi



Sensörler, sosyal medya, uydular, akıllı sayaçlar, mobil cihazlar… Saymakla bitiremeyeceğimiz kaynaklardan her geçen saniye çok büyük bir hızla veri artmaya devam ediyor. Analitiği yapılıp bilgi üretmek için kullanılmadıkları sürece bu büyük veriden büsbüyük dijital çöplükler oluşacak. Bu gelişim çok hızlı gerçekleşmesine karşın veri analitiği konusunda henüz bir standart oluşmuş değil. Büyük veri analitiği konusunda temel bazı başlıkların çok net ortaya konarak iyi anlaşılması gerekiyor. Bu yazıda kaynaktan hedefe bir yolculuk olarak verinin rotası, bu rotada verinin uğradığı duraklar, verinin analitiğinde kullanılan teknolojiler ve onların birleştirilmesini konuşacağız.

Büyük Veri Nedir?

Önce belki birden fazla kez duymuş olabileceğiniz büyük veri'nin 5V tanımı ile sizleri tanıştırmak isterim. Bu 5V'nin her biri İngilizce V harfi ile başlayan 5 kelimeden geliyor. Burada müsaadenizle orjinaline sadık kalarak akılda kalıcı bir tarif yapabilmek adına İngilizce kelimelerle sınıflandırmayı yapacağım.

Birinci V'miz *Variety* (Çeşitlilik). Yukarıda bahsettiğim çok çeşitli kaynaklardan aynı zengin çeşitlilikte gelen veridir, eskiden tablolar halinde tuttuğumuz metin ve sayılardan ibaret verilerle hareket ederken şimdi metin ve sayıların yanına görseller, videolar, ses kayıtları, haberleşme logları, coğrafi konum bilgileri vb. de eklendi.

İkinci V'miz *Velocity* (Hız): Verilerin üretilme ve değişme hızı olarak düşünebiliriz, her şey çok hızlı değişiyor, bu süratle de baş edebilmek gerekiyor. NTV'de yer alan 2017 Ocak rakamlarına göre, sadece internette 1 saniyede; 7 bin 466 tweet atılıyor, Instagram'a 759 fotoğraf ya da video yükleniyor, Tumblr'da 195 paylaşım, Skype'da 2 bin 410 görüşme başlatılıyor, 58 bin 75 Google aramasına 67 bin 666 kez izlenen Youtube videoları eşlik ediyor ve 2 milyon 553 bin 258 elektronik posta gönderiliyor. Saniyede 40 bin 906 GB internet kullanımı gerçekleşiyor.

Üçüncü V ise *Volume* (Verinin hacim olarak büyüklüğü): İnsanlık tarihinde başından beri üretilen tüm verinin yüzde doksanı son 2 yılda üretildi. Bugün İstanbul'dan kalkıp örneğin ABD'nin New York şehrine giden bir yolcu uçağının sadece sensörleri aracılığı ile üretilen veri miktarı onlarca gigabyte ile ifade edilmekte.

Dördüncü V'miz *Verification* (Doğrulama): Güvenlik ayarlarının da değişen verinin hızında kontrol edilerek akış sırasında verinin doğru ve kesin olmasının sağlanmasına bu adı veriyoruz.

Beşinci ve son V'miz ise *Value* (Değer): Üretim, tanıtım ve reklam, pazarlama ve satış, müşteri ilişkileri yönetimi aşamalarında şirketlere kattığı değer. Büyük veriye yatırım yaparak işleyen firmaların kârlarını ciddi bir oranda artırdıklarını McKinsey'in yaptığı çalışmadan öğreniyoruz.

Özetle; hızlı bir artışa sahip, genelde yapısal olmayan (tek bir yapısı olan sayısal, metin, ya da resim olmayıp, birden fazla yapıya sahip

verinin bir arada işlenmesi durumu), büyük hacimlere sahip ve doğru teknikler ile işlendiği vakit bir ekonomik değere dönüşmesi çok büyük bir ihtimal dâhilinde olan veriye "Büyük Veri" adını veriyoruz.

Büyük veri analitiğini ise üç temel kategoriye ayırmak mümkün: zaman içinde biriken verinin (*historical data*) analitiğinin yapılması, gerçek zamanlı olarak verinin analitiğinin yapılması (yani, veri üretilir üretilmez hemen analitiğinin yapılması durumu) ve gerçek zamana çok yakın bir zaman aralığında (*near real-time*) verinin analitiğinin yapılması. Doğru bir kategoride ilerleyebilmek için, veri yolculuğunun rotasının çok iyi tanımlanması ve her bir rotada kullanılacak teknik ve araçların iyi bilinmesi gerekir.

Geçmiş Veri ve Gerçek Zamanlı Veri Ne Demek?

Bu kavramları yolculuğumuz boyunca sıkça kullanacağımızdan biraz daha açıklamakta fayda var. Bugüne kadarki su faturalarınız, sayaç okumaları ve fiyatlandırması geçmiş veri. Olan olmuş, su harcanmış, harcanan miktar dönemsel olarak belli, buna göre faturalandırılmış, hatta faturası otomatik ödeme talimatı ile çoktan ödenmiş bile!

Öte yandan evde şu an banyoda büyük çocuğunuz duş alırken, eşiniz bahçe suluyor, küçük çocuğunuz musluktan suluboya kabına su dolduruyor ve birisi çıkıp sayacın başında durmaksızın hareket eden rakamları gözlemliyor, fiyat bilgisi her an değişiyor çünkü bir sonraki limite erişildiğinde birim fiyat da değişecek, yarış seyreder gibi heyecanlı bir değişim, neler olacak? İşte bu gerçek zamanlı veri ve sürekli değişim halindeyken analitiğinin yapılıp bilgiye dönüştürülmesi bu bilginin de gerek iş zekâsı olarak gerekse ani karar verme gerektiren operasyonel projelerde kullanılması paha biçilmez değerde bir eylem. Musluklar kapanır kapanmaz okunan değerle yapılan hesapları da gerçek zamana yakın olarak yapılan veri analitiği olarak düşünebiliriz.

Bu basit, evsel analojiden sonra, her bir analitik yöntemin kendine özgü farklı teknolojiler ve farklı bir yolculuk rotası içerdiğini söylememiz gerekir.

Geçmiş büyük veri herhangi bir NoSQL veri tabanında veya veri kümeleri halinde ve dosya sistemi üzerinde depolanmış olabilir. SQL veri tabanlarını yapısal yani metin ve sayılardan ibaret olarak tanımlamaya çalıştığımız verileri saklamakta kullanıyorken NoSQL veri tabanlarında adından da kolayca tahmin edebileceğiniz üzere yapısal olmayan veriyi depolamakta kullanıyoruz.

Büyük veri analitiğinde kullanılan birçok teknoloji açık kaynak kodlu projeler olmasının yanında, birçoğu da ticari proje kapsamında Oracle, Cloudera, Microsoft, IBM vb. servis ve altyapı sağlayıcıları tarafından kullanıcılara hizmet olarak da sunulmakta. Ticari çözümler açık kaynak kodlu çözümlerin üzerine geliştirilen ürünler olduğundan, aslında hemen hepsinin temelinde açık kaynak kodlu projeler bulunmaktadır diyebiliriz. Büyük veri analitiğinde herhangi bir uluslararası standart bulunmadığından, yine açık kaynak kodlu bir proje olan Apache Vakfı'nın Hadoop projesi bu anlamda firmalar tarafından kabul görmüş de-fakto bir standart haline geldi.

Oysa Hadoop sadece geçmiş veri (*historical data*) analitiğinde kullanıldığından ve bugün artık büyük veri analitiği sadece geçmiş verinin analitiğinden ibaret olmadığından, büyük veri analitiğinde henüz olgunlaşmayan standartlaşmaya gidebilmek için genel bir referans model tanımlamaya ihtiyacımız var.

Nedir Bu Meşhur Hadoop?

Hadoop'u daha iyi anlayabilmek için biraz teknik detaya girmek gerekiyor. Hadoop, Apache tarafından geçmiş verinin analitiğini gerçekleştirmek amacı ile hayata geçirilen açık kaynak kodlu bir proje. Node (düğüm) adı verilen ve aynı işi yapmak üzere görevlendirilebilecek olan birçok bilgisayarın bir araya getirilmesinden oluşan ve dağıtık mimariye sahip olan bir topolojisi var. Söz konusu topoloji, çok kolay bir şekilde yeni düğümleri eklemeye olanak sağlayan yatay olarak ölçeklenebilir özelliğe sahip. Hadoop'un, HDFS (Hadoop Dağıtık Dosya Sistemi) ve MapReduce programlama modeli olmak üzere iki önemli bileşeni var. Topolojinin üyesi olarak görev alan bilgisayarlara farklı görevler veriliyor, operasyonun yöneticisi gibi düşünebileceğiniz bilgisayara Kontrol/Efendi düğümü deniyor ve diğer bilgisayarlara dağıtılan görevlerin takibini bu düğüm üstleniyor. Kovandaki kraliçe arı ve işçi arılar misali herkes görev yerlerine dağılıp bal yapma işi sona erdiğinde bütün balı bir petekte toplama görevini de yine bu ana düğüm bilgisayarı üstleniyor. Her bir işçi arının yani çalışma düğümü bilgisayarların her seferde ne kadar bal taşıyabileceğini, yani veri işleyebileceğini donanımın gücüne bağlı kepçe (*chunk*) büyüklüğü ile belirleyebiliyoruz. Bu kapasite ne kadar büyükse tek bir seferde o kadar fazla veriyi işleyebilecek olduğundan Hadoop Dağıtık Dosya sisteminin asıl gücü bu parametreden kaynaklanıyor. MapReduce yazılım kütüphanesi de bu dosya sisteminin ayrıklaşmış yapısını destekler şekilde otomatik olarak paralel şekilde çalışır.

IT dünyasında Hadoop'un kullanım alanları ile ilgili birçok örnek bulunabilir. Elektronik ticarette veri analitiği, Hadoop'un gerçek kullanım alanlarından biri. Örneğin Hadoop ile bir elektronik mağazanın günlük, haftalık ya da aylık periyotlar halinde, en çok görüntülenen ürünlerini raporlamak mümkün.

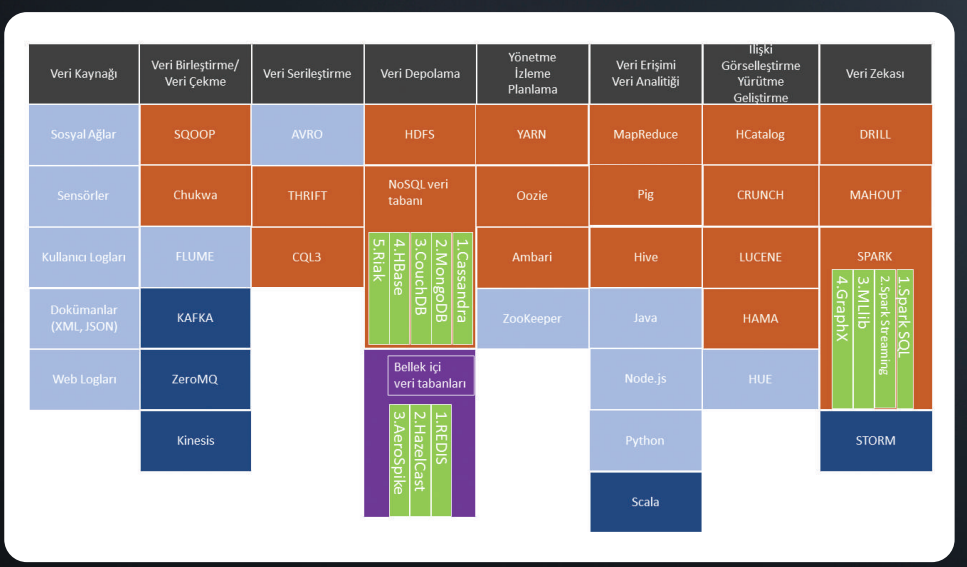
Veri, herhangi bir kaynaktan kendi yolculuğuna başladığında mutlaka bir hedefe gitmek zorunda. Bu hedef ya bir dosya sistemi ya da bir veri tabanı olabilir. Gerçek zamanlı büyük veri analitiği, veriyi daha sonra analitiği yapılmak üzere herhangi bir hedefte depolamadan önce, çeşitli operasyonlarla derhal analitiğinin yapılmasıdır. Bu analitik yöntemi adından da anlaşılacağı gibi çok hızlı bir şekilde verinin doğumuna yakın bir zamanda gerçekleşmekte. Bu analitik tekniği log analizi, risk ve dolandırıcılığın belirlenmesi, güvenlik açısından izinsiz sızma girişimlerinin belirlenmesi, anlık önerme sistemleri (şunu alan bunu da aldı, buna bakan buna da baktı, beğendi vb), iş zekâsı gibi konularda kullanılıyor. Gerçek zamanlı büyük veri analitiği için verinin türüne göre karar verilen ve açık kaynak kodlu proje olan iki tane teknoloji kullanılmaktadır; bunlar Spark yapısının bir kütüphanesi olarak gelen Spark Streaming ve Storm'dur.

Yakın gerçek zamanlı büyük veri analitiği, diğer iki yöntemden biraz daha farklı. Verinin yolculuğunda; veri, analitik işlemi gerçekleşmeden önce depolama tekniği olarak bellek içi veri tabanlarına kayıt edilir. Bu sebeple bellek üzerinde çalışan ve bellek içi veri tabanları olarak adlandırılan depolama ve indeksleme mekanizmaları zorunlu. Günümüzde donanım maliyetlerinin hayli azalması ile bellek içi veri tabanları yaygın olarak kullanılır hale geldi. Bellek içi veri tabanları, NoSQL veri tabanlarında olduğu

gibi dağıtık bir topolojiye sahiptirler. Topolojiye üye olan bütün cihazların belleklerini kullanarak çalışırlar. Yakın gerçek zamanlı veri analitiğinde depolama tekniğinin dışında kullanılan teknolojiler diğer yöntemlerde kullanılanlar ile aynı. Veri doğduğunda kaynağından alınma işlemi, analitik için veri tabanından çekilme işlemi ve analitik sonrası görselleştirme işlemleri her üç yöntemde de ortak duraklar.

dâhil olması gerektiği bilinirse, söz konusu şekle bakılarak hangi teknolojilere ihtiyaç olduğu bulunabilir.

Sonuç olarak büyük veri için tarihsel büyük veri analitiği, gerçek zamana yakın büyük veri analitiği ve gerçek zamanlı büyük veri analitiği olmak üzere üç temel rota mevcuttur. Analitiği yapılacak olan verinin türü ve hangi amaçla kullanılacağı, verinin hangi



Büyük veri analitiğinde en yaygın kullanılan teknolojiler aşağıdaki şekilde sıralanmıştır. Şekil, büyük veri analitiğini genel bir bakış açısı ile özetleyen sekiz temel kategoriden oluşmaktadır. Her kategori kendisi ile ilgili olarak farklı sayıda teknoloji barındırır ve bu teknolojilerin bazıları birbirlerinin yerine kullanılabilen teknolojilerdir. Söz konusu kategoriler bütün veri analitiği yöntemleri için ortaktır, fakat ortak olan bu kategorilerin altında sıralanan teknolojiler her yöntem için farklılık gösterir. Verinin hangi yolculuğa

analitik rotasını takip etmesi gerektiğini belirler. Rotası belirlenen veri kaynaktan doğduğunda, ilgili rotada ilerleyerek ve rotanın her adımında gerçekleşen süreçler ile bir değere dönüşür. Rotanın sonunda elde edilen sonuç ve verinin kendisi durağan hale gelerek sonraki analitik işlemleri için kayıt altına alınır.

Bu analitik işlemler sonucunda eğilimleri, gidişatı anlayıp, geleceği tahmin edebilme yetisi elde edildiğinde hemen her sektör ve organizasyon için şüphe götürmez bir avantaj elde edilmiş olur.

VERİ NE VERİR NE VERMEZ?

Işıl YENİDOĞAN

Khas öğretim görevlisi



Büyük veri, veri madenciliği, veri analizi... içinde bulunduğumuz dönemin en şık, en gösterişli kelimeleri haline geldiler. Aralarında önemli farklar olsa da ortak özellikleri "veriden bilgi üretmek" olan bu kavramlar gerçekten de gördükleri ilgiyi fazlasıyla hakediyorlar. Ancak veri her zaman beklentileri karşılayacak ölçüde konuşur mu? Bazen "ser verip sır vermeyen" veri kümeleri ile karşılaşabilir miyiz? Depolayanından korkması, kalıbından utanması gereken büyük veri kümeleri var mıdır? Yanıtlar sırası ile Hayır-Evet ve Evet.

Büyük Veri Nedir?

Veri analizine veya veri madenciliğine dayalı her proje büyük olasılıkla bir soruya yanıt aramak, bir probleme çözüm bulmak veya bir fırsat yaratmak üzere başlatılır. Bazıları bilimsel merak ürünü olan bu projelerin büyük bölümü de iş zekâsına hizmet etmek üzere planlanmıştır. Bilimin merak ettiği soruların yanıtlarını bulmak, iş dünyasının ihtiyacı olan alışkanlıkları, yönelimleri ortaya çıkarmak üzere çoğu kez ekip ya da ekipler projelerde iş birlikleri içinde çalışırlar.

Bir proje de eğer veri, miktar olarak çok büyük, yapısal olmayan ve zamana bağlı değişen bir yapıdaysa bu proje "büyük veri" projesi adını alır ve bu özelliklerinden dolayı farklı araçlar ve donanımlara ihtiyaç duyar. Yapısal veriyle uğraşıyorsak ve veri tabanlarından bilgi çıkarmaya yönelik bir veri bilimi projesi ise, "veri madenciliği" adını alır. Benzer şekilde yine verinin büyüklüğüne bağlı olarak gelişmiş donanım ihtiyaçları ve yazılım araçları olacaktır. Herhangi bir gizli bilgi peşinde değilsek ve elimizde bulunan veride salt alışkanlıkları, profilleri, eğilimleri görünür kılmaya çalışıyorsak bu kez "veri analizi" projesinde çalışıyoruz diyebiliriz. Adlar değişebilir ancak değişmeyen şey bir "proje konusu" veya "sorusu", bir "ekip" ve "yazılım-donanım" ikilisinden oluşan araçlara ihtiyaç duyulmasıdır. Masada bu üçlü olmadan projeye başlanmaz, bütün oyuncular masadaki yerini almadan da oyun başlamaz.

Projelerde kullanılan araçlar ve yöntemler başka bir yazının konusu. Bu yazıda "veri"nin bir şey vermesi için gerekli insan gücünü, bazen ekibin bütün çeşitlilik ve yeterliliğine rağmen ağzından laf alınamayan veri kümelerini konuşacağız.

Bir veri madencisi her projesinde farklı alanlardan verilerle çalışabileceği gerçeğini baştan idrak ve kabul etmiş olmalıdır. Buluttaki veya lokaldeki veriler bazen finans sektöründen gelebilir, bazen meteoroloji, bazen perakende mağaza zincirleri bazen de tıpla ilgili veriler olabilir. Veri analizi için gerekli algoritmaları, ön hazırlık, veri ön işleme aşamalarını, verinin dönüştürme ihtiyaç ve yöntemlerini çok iyi bilseniz bile ekibinizde alan uzmanı bir müttefikinizin olması cephenizi müthiş kuvvetlendirecektir. Finans projesinde masada bir finansçı, tıbbi veri ile uğraşırken ortamda bir hekim olması paha biçilmez bir değer, her şeyden önce bir ihtiyaçtır. Her projenizde ister istemez projenin ilgili olduğu alana dair fikir sahibi olsanız bile bu bir sonraki projeyi tek başınıza

yürütmenize yeter koşul yaratmaz. Sadece alan uzmanı ile daha rahat iletişim kurmanızı sağlar. Elbette bu aşamada alana dair, özellikle proje konusu ile ilgili birkaç okumanın yaratacağı fayda da tartışılamaz.

Veri Kendinden Bekleneni Bazen Neden Vermez?

Çünkü bazen sorulan soru ile eldeki veri alakasızdır. Örneğin elinizde hatırı sayılır bir miktar öğrenciye ait kayıtlar var ve bu kayıtlar öğrencilerin sadece ad, soyad, adres, telefon ve kaçıncı sınıf olduğu bilgilerini içeriyor olsun. Siz de bu dönem sonunda kayıtlarda yer alan öğrencilerin ne kadarının 1.8 GNO'nun altına düşüp başarısız olarak etiketleneceğini tespit etmek ve belki buna göre birtakım önlemler almak istiyorsunuz. E ama verilerin içinde öğrencilerin başarı durumlarıyla ilgili hiçbir şey yok!? Sorduğunuz sorunun yanıtını bu veri seti içinde bulmak imkânsız! Ne yapmak lazım? Muhtemelen bir başka veri tabanında aynı öğrencilerin bu zamana kadarki not ortalamalarını ve başarı durumlarını tutan bir küme vardır; bu kümeye erişim izni çıkartıp, bu izni alabilmek için gerekirse projeye bu erişim izni olan tüzel veya özel kişiyi de katıp iki kümenin kesişim ve fark kümelerinden hareketle çalışmaya yön vermek lazım. "Ama bu çok bariz, ortada başarıya dair bir veri yokken nasıl bu soru sorulur ki?" deyip gülmeyin, gerçekten başıma geldi, sorulan soru ile alakasız bir veri seti, karşımda projenin sahibi, "bilgisayarınız yok mu yapın işte!" diyor. Yapamam! Yapmayı çok istiyorum inanın ama bunu yapabilmem demek bildiğiniz hayal gücüme dayalı çıkarımdan sonuca gitmem demek. Şu ana kadarki bilgi ve tecrübelerim bana hayal gücünün çok gerekli bir şey olduğunu, her şeyin hayal etmekle başladığını öğretmiş olmakla beraber henüz hiçbir projemin sonucunu da hayal gücüme dayanarak ilan etmedim!

Çok sevdiğim bir Temel fıkrası vardır. Ormanda kaybolan bir fille ilgili. Nice ajanlar girip ormanda fili arıyor ancak bulamıyorlar, bizim Temel üç gün sonra ormandan yanında bir zürafa olduğu halde çıkageliyor, zürafa karşısında gördüğü yetkililere ağlayarak koşarken "abi yemin ediyorum ben filim, Allah aşkına kurtarın beni bunun elinden" diyor. Sonuç olarak ikna kabiliyeti yüksek bir araştırmacı dahi olsanız bazen veri zürafadan daha inatçı çıkabiliyor!

Ya Sadece On Kişi Varsa?

Bazen de son derece alakalı bir veri kümesi ile kutsanmışsınızdır ancak miktar yetersizdir. Böyle taşıma suyla değirmen döndürmeye çalıştığımız projeler de oldu, ancak sonucun güvenilirliği hepimiz tarafından şüpheyle karşılanırken üst yönetimlerin buna dayanarak aksiyon almasını sağlamak artık bizim değil risk yöneticilerinin sorunu haline geldi.

Bu duruma örnek teşkil edecek bir olayı ise bir Tıp Bilişimi Kongresi'nde yaşadım. O vakitler bir tıbbi veri seti üzerinde ve İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi hekimleri ile birlikte çalışıyorduk, güzel bir çalışmamız olmuştu, yayınlamaya değer bulduk ve uluslararası bir tıp bilişimi kongresine gönderdik. Sunum günü geldiğinde kendi oturumumuzda başka yayınları da dinliyorken bir sunumda yapılan çalışmada toplam kayıt sayısının 10 olduğunu gördük. Basımda hata yok, doğru okudunuz, toplam 10 hasta üzerinde çalışma yapılmış gerçekten. Elbette salondaki hemen herkesten benzer sesler yükseldi, bu kadar az kayıtla nasıl? Bu olur, şu olur, çalışmaya katılanların yüzde şu kadarı kadın, şu kadarında iyileşme görüldü... Hepi topu tamamı 10 kişi zaten. Nasıl bu kadar kati rakamlar verilebilir? Yanıt şöyle oldu: Çalışmanın konusu meğer çok spesifik bir kanser türü imiş, boğazın burunla birleşen bölümü yani nazofarenks kanseri ile alakalı bir çalışma, bu o kadar ender görülen bir şeymiş ki yıllar içerisinde kendilerine gelen vakalardan biriktirebildikleri sadece toplam 10 kişi olmuş. Yani günün sonunda çalışmanın sahipleri "100 kişi vardı da biz mi sormadık?" diyecek oldular. İçimizden geçen, keşke o 10 kişi de olmasaydı, kimse kanser olmasaydı duygusu elbette.. Neticede insan kurtarmak için araştırma yapmak zorundalar, ancak araştırma için de ellerinde yeterince örnek yok. Zor bir problem.

Veriden Bilgiye Giden Yolda Bazı Engeller!

Bir sonraki probleme geçelim. Veri toplama yöntemleri arasında anket, müşteri kayıtları, işlemleri (alış-veriş) ve ölçümler (sensörler, ses kayıtları, uydu verileri vb.) yer alır. Anket ile veri toplayarak bilgiye ulaşmaya çalışan meslektaşlarımın sayısı azımsanamayacak kadar fazla çünkü ne yazık ki ülkemizde halen veri paylaşımı ile ilgili bir politikamız yok, bu konuda atılmış bir adım da yok. "Kişisel Verilerin Korunması Kanunu"na ters düşmeden de pekâlâ anonimleştirme algoritmaları kullanılarak ciddi ve gerçek kayıtlar araştırmacılar ile paylaşılabilir ve bu şekilde her sektörde ilgili veri çalışmaları daha yüksek doğrulukla yapılmaya başlanabilir. O gün gelene kadar anket, yine kurtarıcı veri toplama yöntemlerinden biri olacak gibi görünüyor. Buna karşın insanların yalan söyleme alışkanlık ve yetenekleriyle baş edecek bir algoritmamız var mı? Yok. Herhangi bir istihbarat kurumu da değiliz ki ankete katılan kişiyi çapraz sorguya alalım. Elbette anketlerdeki yanıtların doğruluk ve kendi içerisinde tutarlılığını sağlamak için özel soru hazırlama teknikleri kullanıyoruz ancak insan faktörü her zaman ezici üstünlükle ağırlığını hissettiriyor. Kilosunu ve günlük tükettiği şeker miktarını yanlış rakamlarla bildiren kişilerden yola çıkarak diyabetle ilgili güvenilir bir çalışma yapamazsınız. Sanırım bunun tek çözümü farkındalık yaratmaya yönelik çalışmalar ile kişisel bilgilerinin güvende olacağına ilişkin ikna kabiliyeti ve hassasiyet.

Veri kendisine sorulan sorulara yanıtlarını vermemekte direniyorsa bazen bunun nedeni gerçekten de tekniğin yanlış olmasından kaynaklanabilir. Veriler konuyla son derece alakalıdır, üzerinde çalışmaya elverişli bir veri kümesi elde edilmiştir ve doğru olduğuna da inanç tamdır ancak yine sorunlarımız bitmedi. Verinin ilk hazırlanma aşamalarında gözden kaçan bir detay veya yapılan bir hata, veriden bilgi çıkarımı yolculuğunun ileri aşamalarında da omuzlarda gittikçe ağırlaşan bir yük gibi taşınır. En kötüsü de ya sorunun yanıtına ulaşmanıza engel olur ya da yanlış sonuçlar elde etmenize neden olur. İster acemilik ister ehil olmama isterseniz dikkatsizlik deyin, bir veri madenciliği projesinde sarf edilecek eforun neredeyse %70'ine mal olacak veri hazırlığı (aykırı değer analizi, eksik-kayıp değerler, veri dönüşümü) evresinde şu veya bu nedenle hata yapmak projenin direkt başarısızlık sebebi olabilir. Bu aşamada dikkatli olmayı ve bazen teyit için birden fazla kontrol yapmayı önerebilirim. Proje körlüğü denilen durumdan sakınmak için de bazen ekranınıza ekipten bir başkasının bakmasını rica etmelisiniz. Sizin göremediğinizi o iki dakikada görüp dikkatleri oraya çekebilir.

Veriden bilgi üretmeye giden heyecan verici yolculukta aşılması gereken engellerden sadece birkaç tanesini özetlemeye çalıştım. Engeller bazen gerçekten zorlayıcı olsalar da bilgi keşfinin verdiği heyecan ve haz harcanan bütün eforu katlanılır kılıyor.

BÜYÜK VERİ VE ENFORMASYON SİYASETİ

H. Akın ÜNVER

Khas öğretim üyesi

Son yıllarda popülerleşen "büyük veri" kavramı, tanımı ve uygulama alanları itibariyle henüz genel olarak kabul edilen bir kavramsallaştırma safhasına erişememiş durumda. Mevcut dijital jargonda, büyük veri üç ana süreç ekseninde tanımlanıyor: Geleneksel veri işleme tekniklerinin yetersiz kalacağı büyüklükte ve karmaşıklıkta (çoğunlukla gerçek zamanlı) veri; ekseriyetle insan ilişkileri, tercihleri ve davranışlarına ait bilgi içeren veri ve hem yapılandırılmış, hem de yapılandırılmamış (ham) veri. Bu denli büyük ve karmaşık insan odaklı veri gruplarının yeni teknik ve araçlarla işlenmesi, siyaset, uluslararası ilişkiler ve temsil konusunda çok büyük değişiklikler getirmekte. Bu değişikliklerin büyük bir bölümü toplum ve insan için faydalı olsa da (ekonomik ihtiyaçların tetkiki, sağlık/kamu hizmetleri, asayiş ve emniyet süreçleri v.b. gibi), büyük veri devriminin getirdiği bazı diğer yenilikler de olumsuz veya öngörülemez olmakta.

Günümüzde toplumlar, akıllı telefonlar, bilgisayarlar ve internet-bağlantılı araçlar üzerinden daha önce eşi benzeri görülmemiş büyüklük ve detayda dijital veri üretmektedir. Seçmen, sağlık, banka



kayıtları, internet alışverişi gibi bir dizi dijital veri toplama platformunda sadece binlerce kişinin değil, bütün toplumların verileri kaydedilmekte, bu denli büyük verilerin işlenmesi ve depolanması da dijital meşruiyet, egemenlik ve milliyetçilik gibi kavramları siyasi tartışmaların merkezine oturtmaktadır. Örneğin Venezüela'da faal olan bir Fransız sigorta kuruluşu, sigorta sattığı kişilerin sağlık, adres, banka ve kişisel bilgilerini nerede depolamalıdır? Bu veriler Venezüela'da siber alanda çok iyi korunmayan ulusal veri işleme birimlerinde mi, Fransa'da daha iyi korunan, ama fiziksel olarak Venezüela'da bulunmayan depolama tesislerinde mi, yoksa üçüncü bir ülkede mi muhafaza edilmelidir? Şayet bu bilgiler üçüncü bir ülkede depolanacaksa, bu veriler hukuken Venezüela'nın mı, Fransa'nın mı, yoksa üçüncü ülkenin mülkiyetinde midir? Her ne kadar bu sorular hukuk soruları gibi dursa da aslında siyasi meşruiyet-egemenlik ve küresel dijital diplomasi-istihbarat meseleleridir.

Akademik literatürde devam eden siyaset-büyük veri tartışması iki alanda gruplanmaktadır. Bu analitik gruptan ilki, büyük veri uygulamalarının sosyal ve siyasal sorunları ne şekilde çözebileceği ve vatandaşların dijital dünya üzerinden ne şekilde daha bilinçli, aktif ve verimli siyasal öznelere dönüşebileceğini kavramsallaştırmaktadır. Bu bağlamda büyük veri etiği, algoritmaların sosyal-siyasal değişimlere karşı nasıl daha hassas hale getirilebileceği gibi konular istatistiksel ve bilişimsel metotların öncelikli ilgi alanıdır. Büyük veri, enformasyon etiği ve dijital psikoloji gibi yeni gelişen alanlar da bu açıdan son derece önemlidir. Akademik literatürdeki ikinci önemli yaklaşım da bir özne olarak verinin kendisine dönük tartışmaları içerir: veri üretimi, işlemesi ve depolanması gibi konuların siyaset, diplomasi, psikoloji ve hukuk gibi alanlara etkileri bu bağlamda çalışılmaktadır. Aynı şekilde, "eleştirel algoritma çalışmaları" olarak yeni beliren bir alt çalışma disiplini de "objektif" olarak kabul edilen otomasyon ve programlama süreçlerindeki gizli önyargılar ve sübjektiviteler üzerine yoğunlaşmaktadır. Aynı şekilde "eleştirel makine öğrenmesi çalışmaları" (*critical machine learning studies*) da otomasyonun farklı süreçlerinin insan ve siyaset üzerindeki etkilerini kavramsallaştırır. Bu konuda DARPA'nın Celalabad projesi ve Black Sage Technologies'in hava alanlarının korunmasına dönük projeleri iki kısa ve önemli örnek olarak gösterilebilir.

ABD'nin Afganistan'da başlattığı "veri karşılığı bira" programı, büyük verinin stratejik ve askeri olarak ne şekilde kullanılabileceği konusunda önemli bir ilk vakadır. Amerikan İleri Savunma Araştırma Projeleri Ajansı (DARPA), 2009'da Afganistan'da kitle-kaynaklı (*crowdsourced*) bir istihbarat projesi başlattı. Bu projenin amacı bölgede faal olan ve her ülkeden gelen insani yardım çalışanları, özel güvenlik taşeronları ve inşaat firmaları işçilerinin, çalıştıkları bölgeyle ilgili insani, iktisadi ve operasyonel bilgileri tek bir internet kafede (Celalabad'daki Taj Mahal konukevi) toplamaları, bunun karşılığında da bedava birayla ödüllendirilmeleriydi. İlk başta tek tük toplanan veriler, birkaç ay içerisinde *terabytelar* dolusu görsel, video ve yazılı enformasyonla birlikte geniş bir büyük veri istihbarat platformuna dönmüştü. Bu büyüklükte bir veri akışı, Amerikan veya NATO ordularının kullandıkları Nexus 7 sisteminden bile daha büyük ve yerel bilgileri sisteme aktardığından, Amerikan ordusu da çok kısa bir süre sonra bu kitlekaynaklı açık istihbarat sistemine geçti. Artık yeni istihbarat rejiminde gizlilik değil, açıklık önemliydi ve enformasyonu bulmak kadar paylaşmak da istihbarata katma değer sağlayan bir faktördü.

Ancak kitle kaynaklı istihbarat sadece bir başlangıç noktasıydı. Büyük veri tartışmaları hızla makine öğrenmesi (*machine learning*) gelişmeleri eksenine girerek otomatik veri toplama, gerçek zamanlı veri işleme ve istihbarat içeren bilginin ne olduğuna dair kararın bilgisayarlar tarafından verilmesi gibi konuları içermeye başlamıştır. Örneğin Amerikan Black Sage Technologies şirketi, bir havaalanında bulunan 50 kameranın aynı anda yüz tanıma algoritmaları yoluyla, daha önce sisteme yüklenen Interpol veritabanında aranan kişileri bulabilecek yeterliliğe gelebileceği bir proje başlatmıştır. Aynı şirket, ABD'deki bütün havaalanlarını, eşzamanlı uydu tetkik sistemi yoluyla, tehlikeli mesafede yaklaşan insansız hava araçlarına karşı otomatik olarak uyarma ve koruma sistemini de test etmiştir. Her iki sistem de makine öğrenmesi yoluyla havaalanlarına karşı yapılabilecek terör saldırılarını öngörme, modelleme ve engelleme gibi konularda öncü kabul edilmektedir.

Bu gelişmeler gözetim, denetleme ve mahremiyet gibi konuları da ilgilendirmektedir. Privacy International örgütünün bir araştırmasına göre dünyada bugün 80 milyondan fazla izleme kamerası bulunmaktadır. Bu kameraların sadece %6'sı polis ve özel güvenlik elemanı gibi gerçek kişiler tarafından takip edilmekte, kameraların geri kalanı yalnızca kayıt almaktadır. Tartışmanın güvenlik tarafından gelen kişiler, bu kameraların ancak bir suç işlendikten sonra hukuki delil olarak kullanılabildiğini, ancak öngörü veya gerçek zamanlı suç engelleme konularında yetersiz kaldığını iddia etmektedir. Bu görüşe göre gözetleme kameraları makine öğrenmesi yoluyla suçu tanıyan ve güvenlik birimlerini uyaran bir merkezi otomasyon sistemine bağlanmalı ve bu yolla gerçek kişiler tarafından gözetlenmeyen kameralar da suç teşhis ve engellenmesine dönük kullanılmalıdır. Tartışmanın öbür tarafında ise mahremiyet konusu yer almaktadır. Suç engelleme isteği öne sürülerek kişisel bilgilerin ve davranışların gerçek zamanlı olarak analizi, mahremiyet konularına dönük bir dizi etik problemi de beraberinde getirmektedir. Bu problemlerin başında sürekli gözetim, kişisel bilgilerin korunması ve bu gözetim teknolojilerinin ne gibi bir hukuksal çerçeve içerisinde kullanılacağı gelmektedir.

Sonuç olarak büyük verinin siyasi boyutu, hâlâ büyük tartışmaların devam ettiği ve etik kurallarının sürekli tekrardan belirlendiği bir alan olmaya devam edecektir. Büyük veri, algoritma ve otomasyon konuları, öngörülebilir bir süre içerisinde son derece büyük tartışmalara ve sosyal değişimlere gebedir. Büyük veri gelişmelerini destekleyenler, yeni teknolojilerin, tarihte hiçbir zaman olmadığı kadar geniş ve detaylı sosyal veri üretmesi sebebiyle olumlu olduğunu ve toplumsal sorunların çözümünde bu verilerin çok önemli olduğunu önermeye devam edecektir. Bu teknolojik devrime karşı daha temkinli yaklaşanlarsa veri toplama süreçlerinin etik, mahremiyet ve hürriyetler ekseninde devam eden büyük sorunlarına dikkat çekmektedir. Her ne kadar büyük veri ve dijital teknolojiler dünya çapında tartışmaların ortasında yer alsa da, mevcut teknolojik devrim duraksamadan, daha da büyük bir hızla gelişerek mevcut siyasal ve toplumsal kabulleri temelinden değiştirecektir. Bu değişimlerin daha insan odaklı ve yapıcı gelişmesi için bilgisayar mühendisliği içerisinde yer alan programlama, otomasyon ve algoritma gibi konuların giderek daha fazla sosyal bilimler alanına kayması ve objektif olarak kabul edilegelen bu konuların, sosyo-kültürel sübjektivitelerinin daha net bir şekilde kavramsallaşması gerekmektedir.

DEVRİMCİ TEKNOLOJİLERİN ARDINDAKİ GÜÇ: AĞ BİLİMİ

Mehmet AYDIN – N. Ziya PERDAHÇI

Khas öğretim üyesi - Mimar Sinan Güzel Sanatlar Üniversitesi öğretim üyesi



Çevrimiçi sosyal ağ konusuna sanayi ve akademide ilginin giderek arttığı ve uygulama alanlarının giderek genişlediği görülmekte. Bu genişlemenin arkasındaki temel, ağ bilimi (*network science*) adıyla bilinen yeni bir eklektik disiplin. Bu makale, oldukça genç ve farklı bir bilim dalı olan ağ biliminin yaklaşımını takdim ederek daha iyi anlaşılmasını amaçlamakta.

Hayatımız, sosyal toplumdan cep telefonu ağlarına kadar pek çok sistemle çevrelenmiş durumda. Sosyal bir toplumun sağlıklı (kesintisiz) işleyişi milyarlarca bireyin işbirliğine bağlı; tıpkı milyarlarca cep telefonunu bilgisayar ve uydularla bütünleştiren iletişim altyapısının işleyişi gibi. Etrafımızda olup biteni muhakeme ve idrak edebilme becerimiz beynimizdeki milyarlarca sinir hücresinin uyum içinde işleyişi ile mümkün. Tüm bu sistemlere karmaşık (*complex*) sistemler adı veriliyor (Lorenz, 1963; Mandelbrot, 1982). Gerek günlük yaşantımız içinde gerekse bilimde ve ekonomide sahip oldukları yadsınamaz rolleri, karmaşık sistemleri anlama, matematiksel tanımlarını yapma, tahmin ve nihayetinde kontrol etme problemlerini 21. yüzyılın en önemli zihinsel ve bilimsel meydan okuyuşlarından biri kılmıştır.

Easley ve Kleinberg, ağ biliminin eklektik çoklu bir bilim dalı olarak araştırmacılar için yeni araştırma problemleri veya mevcut araştırma problemlerine yeni yaklaşımlar sunmasıyla öne çıktığını vurgulamıştır. Diğer yandan bu bilimsel eklektik yapının bilgi kuramı (*epistemology*) ve varlık bilimi (*ontology*) açısından dikkatle inşa edilmesi gereklidir. Ağ bilimi alanındaki gelişme daha sonra detaylandırılacağı üzere özellikle bilgi teknolojilerindeki gelişmeye bağlı olarak hızlı ve farklı araştırma alanlarında parçalı (Dunn ve Westbrook'un çalışmalarında ölçüt ve sonuçların karşılaştırılması problemi örnek teşkil eder) şekilde gelişmektedir. Bu bağlamda ağ biliminin daha iyi tanıtılması ve ortaya koyduğu yaklaşımın özellikle bu alana ilgi duyanlar açısından daha iyi anlaşılması gerekmektedir.

Ağ bilimi konusunda ülkemizde yapılan akademik çalışmaların bazı istisnalar dışında salt farkındalık seviyesinde olduğu düşünülürse, Kadir Has Üniversitesi Ağ Bilimi odak alanı çalışmalarımız Türkiye'de ağ bilimi genelinde ve sosyal ağlar özelinde bilimsel çalışmaların gelişimine katkı sunmaktadır. Ağ bilimi alanında akademik bilgi ve becerenin kazanılması ve bu kapasitenin sahada etkin kullanımı ancak sıkı bir üniversite-sanayi işbirliği sayesinde gelişebilir kanısındayız. Çalışmalarımızın sanayi ve akademiye müşterek fırsatlar yaratacağı düşüncesinden hareketle veri üreten ve kullanan sektörlerle sürekli temas halinde bulunuyoruz.

10 Haziran 2015 tarihinde Ulusal Ağ Bilimi Çalıştayı Kadir Has Üniversitesi'nde çok sayıda akademisyen ve sektör temsilcilerinin katılımıyla gerçekleşti. Detaylı bilgi için agbil15.khas.edu.tr adresini ziyaret edebilirsiniz. Çalıştay, üniversitenin Teknoloji ve Toplum Uygulama ve Araştırma Merkezi ve Mühendislik ve Doğa Bilimleri Fakültesi desteğiyle organize edildi. Çalıştay gündemi, değerli Bilim Kurulu üyelerinin değerlendirmeleri sonucunda bildiri sunumları, davetli konuşmacılar ve panelden oluşmuştu. Çalıştay sonrası ayrı bir değerlendirmeyle seçilen çalışmaları içeren kitabımız ilgilenenler için Kadir Has Üniversitesi Teknoloji ve Toplum Araştırma ve Uygulama Merkezi yayınevi tarafından basılmıştır.

Çalıştayda sunulan çalışmalar kuramsal ve ampirik (gerçek hayata ait veriler) türde ve çok farklı uygulama alanlarından gelmekteydi. Seçmenlerin fikir birliğine nasıl bir etkileşim sonucunda varabileceği, sihirli tat formülü hipoteziyle dijital gastronominin doğuşu, diplomasideki haberleşmenin görünmeyen yüzü, sağlık bilgisi paylaşımında sosyal medyanın rolü ve güvenli kaynak arayışı, Türkiye internetinin kritik oyuncuları ve ilişkileri Çalıştay'da ortaya konulan değerli çalışmalara birkaç örnek.



Kadir Has Üniversitesi'nde Veri-Ağ Bilimi Temalı Çalışmalar

Sağlık konusunda sosyal medyanın rolü ve güvenilir bilgi kaynağı arayışı: Sağlık hizmetlerinin özelleştirilmesiyle sağlık sorunları konusunda bilgi paylaşımı ve doğru doktor seçimi önemli bir mesele olmaya başlamıştır. Sosyal medya ve web sayfaları her geçen gün hem sağlık sorunlarına çözüm arayanların hem de sağlık hizmeti sunan kurumların ilgisini çekmeye başlıyor. Doğru ve tam bir sağlık bilgisine ulaşma bağlamında, güvenilir bilgi kaynaklarına nasıl ulaşılır, doktorlar nasıl tercih edilir, bilgi arayan-bilgi sağlayan etkileşiminin görünmeyen yüzü nedir vb. sorulara etkileşimlerin ardındaki işleyişi karmaşık sistem bakış açısıyla analiz ederek çok sürpriz cevaplar ortaya çıkarıyoruz. Burada gün içinde yüzbinlerce doktor-hasta etkileşiminden bahsediyor, kimlerin nasıl yönlendirici rolüne sahip olduğunu, nasıl topluluklar oluştuğunu ortaya çıkarıyoruz.

Diplomasinin Görünmeyen Ağları: Wikileaks verisiyle yüzbinlerce telgrafa ait çok detaylı bir biçimde veri hazırlama sonrasında ağ bilimi yaklaşımıyla ilişkisel analizi yapılmaktadır. Haberleşmede hangi diplomatik unsurların (konsolosluklar, büyükelçilikler, askeri örgütler, vb.) öne çıktığını, kimlerin haberleşmede merkezi rol aldığını, hangi ülkedeki diplomatik unsurların bazı diplomatik yazışmalardan özellikle dışlandığını ve zamana bağlı olarak nasıl bir yönelim olduğunu anlamaya çalışıyoruz.

Ağ Biliminin Gelişimi ve Temel Özellikleri

Değerli bilim insanları Albert, Barabási ve Newman'ın öncü çalışmaları, 21. yüzyılın başlarında ağ biliminin ortaya çıkışı, bilimin yukarıda bahsedilen meydan okuyuşa cevap verebileceğinin canlı göstergesidir. Aslında, her bir karmaşık sistemin bünyesinde, sistem bileşenlerinin etkileşimlerini kodlayan karmakarışık bir ağ vardır: Hangi iletişim cihazlarının birbirleriyle etkileşim içinde olacağını tanımlayan iletişim ağı, modern bir haberleşme sisteminin kalbidir; ticari ağların mal ve hizmetlerin değiş tokuşunda oynadığı kilit rol bir diğer örnektir.

21. yüzyılın önde gelen devrimci teknolojilerinin ardında esasen ağlar mevcuttur. Google'dan Facebook'a, Cisco'dan Twitter'a kadar pek çok teknoloji şirketinin gerçek gücü sahip oldukları ağlardır. Sözün özü: ağlar bilime, teknolojiye ve doğaya "sıradan" bir bilimsel incelemeye göre çok daha yüksek bir mertebede nüfuz etmemize imkân verir. Sonuç olarak, günümüzde yaygın kanı şudur ki, karmaşık sistemleri modellemek ve incelemek, temellerinde yatan ağları derinlemesine anlamaktan geçer.

21. yüzyılın ilk on yılında ağ biliminde yaşanan hızlı gelişimin temelinde yatan esas neden, ilk bakışta birbirlerinden tamamen farklı görünen ağların oluşum ve gelişim mekanizmalarının bir takım ortak temel yasalarla izah edilebileceğinin Feigenbaum, Albert ve Barabási tarafından keşfine dayanır. Yani, doğada, sosyal toplumlarda ve teknolojide gözlemlenen ağların pek çoğu büyüklüğü, biçimi, doğası ve kapsamından bağımsız olarak ortak örgütsel ilkelerle sürdürülür. Bir başka deyişle, yine öncü çalışmalarıyla tanınan Watts, Strogatz, Wilson, Witten, Sander'in çeşitli ifadelerle vurguladığı gibi, ağlar yeterince soyut bir seviyede (matematiksel olarak) ele alındıklarında aralarındaki benzerlikler farklılıklardan çok daha fazladır.

Matematiğin oldukça üretken bir alanı olan ve seçkin matematikçi Leonard Euler'in 18. yüzyılda "Königsberg'in Yedi Köprüsü" problemini çözmesiyle doğduğu kabul edilen çizge kuramı (*graph theory*), soğuk savaş yıllarında sıra dışı üretkenliği ile meşhur Paul Erdös ve çalışma arkadaşı Alfréd Rényi'nin ağ bilimine temel teşkil eden rassal ağlar konusunda gerçekleştirdiği değerli çalışmalar ve en büyük ağlardan biri olan internetin kırk yaşını çoktan geride bıraktığı düşünülürse, ağ biliminin ortaya çıkmakta neden bu kadar geciktiği akla takılabilir. 21. yüzyılın başlarında ortaya çıkan özel bir durumun yarattığı katalitik etki, ağ biliminin yeni bir disiplin olarak ortaya çıkmasına neden olmuştur: Bu özel durum, ağ haritalarının ortaya çıkışıdır. Milyonlarca hatta milyarlarca bağlantıdan oluşan bir ağın incelenebilmesi için öncelikle sisteme ait bir bağlantı haritasına ihtiyaç vardır. Örneğin sosyal bir ağın haritası bireyin arkadaşlarını, bireyin arkadaşlarının arkadaşlarını vs. ve tüm bireyler arasındaki muhtemel arkadaşlık ağını içerirken, www'in (*world wide web*) haritası web sayfalarını ve sayfaların birbirleriyle yaptıkları bağlantıları içerecektir. Bilgisayar teknolojilerinde yaşanan hızlı gelişim, nihayet ağ verilerinin saklanması, tasnif edilmesi ve ağ haritalarının oluşturulmasına olanak vermiştir. Özellikle 1990'ların sonlarında harita çıkartma çalışmaları ivmelenmiştir. Öne çıkan örnekler arasında, internet ağının haritasını yapan CAIDA ve DIMES projeleri, biyologların insan hücrelerindeki proteinlerin birbirleriyle kurdukları bağlantıların haritasını çıkartmaya yönelik milyonlarca ABD dolarlık projeleri ve memeli canlıların sinir ağlarının haritasını yapmaya yönelik Human Connectome projesi sayılabilir.



Aslında pek çok ağ haritası başlangıçta ağ fikri düşünülmeksizin inşa edilen veri tabanlarından, veri ön işleme yazılımlarıyla ve hatırı sayılır insan gücü harcanarak üretilmektedir. Sinema oyuncularıyla, rol aldıkları filmlerin veri tabanı olan IMDB.com'dan elde edilen haritalar buna tipik bir örnektir. Ağ bilimini, hem kapsamı hem de yöntembilimi ayrıcalıklı kılar. Ağ bilimi doğası gereği disiplinlerarası, deneysel (veriye dayalı), nicel ve matematiksel, aynı zamanda hesaplamalı bir bilimdir. Genç bir bilim dalı olmasına rağmen ağ biliminin etki alanı geniştir.

Ekonomi: 21. yüzyılın önde gelen teknoloji şirketlerinin pek çoğunun (Google, Facebook, Cisco, Apple, Akamai) hem iş modeli hem de teknolojisinin temel bileşeni ağdır. Google, işletme anlamında dünyanın en büyük ağ haritası çıkartan firmasıdır. Google'ın site sıralamasında kullandığı PageRank algoritması tam anlamıyla bir ağ bilimi ürünüdür. Ağ biliminin önemi Facebook'un yükselişiyle hızlanmıştır. Facebook'un iddiası tüm dünyanın sosyal ağ haritasını çıkarmaktır. Yükselen değer sadece Facebook değildir. Twitter'dan Orkut'a kadar pek çok sosyal medya aracının aktif kullanıcı sayısı sürekli artmaktadır. Tüm bu sitelerin ekonomik kazanımlarının altındaki ateşi besleyen yakıt ağ bilimidir. Arkadaş tavsiyesinden kişiye özel reklam üretmeye kadar tüm araçlar ağ bilimi sayesinde geliştirilmektedir.

Sağlık: İlaç tasarımından metabolizma mühendisliğine kadar, insan genomu projesi 2001'de tamamlandığında insan genlerinin tam listesi belirlenmiştir. Projenin ardından hücrelerimizin nasıl çalıştığını ve hastalıkların kaynağını açıklayabilmemiz için genlerin etkileşim haritasının çıkartılması gerekmiş ve hücresel ağları araştıran ağ biyolojisi ve hastalıkları araştıran ağ ilaç bilimi ortaya çıkmıştır. Ağ farmakolojisinin temel amacı yan etkisiz ilaçlar üretilmesidir. GeneCo ve Genomatica, farmakoloji sektöründe ağ bilimine dayalı Ar-Ge firmalarına örnektir.

Güvenlik ve terörizmle mücadele: Terörist eylemlerin dizginlenmesinde ağ bilimi yaklaşımı giderek yaygınlaşmaktadır. Körfez Savaşı'nın ardından Saddam Hüseyin'in ele geçirilmesi ve 11 Mart 2004 günü Madrid trenine yapılan terörist saldırının faillerinin yakalanması ağ biliminin bazı önemli başarılarına örnek teşkil eder. Ağ bilimi kavramları askeri doktrini de etkilemiş, ağ savaşları başlamıştır. 2009 yılında ABD savunma bakanlığı güvenlik alanında yapılan ağ bilimi Ar-Ge çalışmalarına 300 milyon ABD doları destek sağlamıştır.

Salgın hastalıklarla mücadele: Salgın tahmininden ölümcül virüslerin önlenmesine kadar yaygın etkisi olduğu bilinmektedir. Örneğin, N1H1 (domuz gribi) vakası, virüs salgınları tarihinde bir ilke sahne olmuştur; salgının kesin olarak nasıl yayılacağı tepe noktasına erişmeden aylarca önce Balcan ve arkadaşlarının ağ bilimine katkı sunan çalışmaları ile kestirilmiştir. Halen salgın tahminleri ağ biliminin en yaygın kullanıldığı alanlardan biridir. Bir başka çarpıcı örnek, Wang ve arkadaşlarının 2010 yılında cep telefonlarının salgın bir hastalığa neden olabilmesine yol açacak senaryoları ağ bilimi yaklaşımı ile tespit etmeyi başarmalarıdır. Yine Balcan ve arkadaşlarının diğer bir çalışmasında, aynı senenin sonlarına doğru Çin'de günde üç yüz binden fazla cep telefonuyla yayılan salgın bir hastalığın gelişiminin araştırmada belirlenen senaryolara aynen uyduğu görülmüştür.

Beyin araştırmaları / Sinir ağının haritasını çıkartmak: İnsan beyninde milyarlarca sinir hücresinin bağlantılarından oluşan sinir ağı, ağ bilimiyle en az incelenmiş alanlardan biridir. Bunun yegâne nedeni sinir ağlarının haritasının henüz çıkartılamamış olmasıdır. Sadece 300 sinir hücresine sahip olan Caenorhabditis elegans (yuvarlak solucan), sinir ağı haritası oluşturulabilmiş yegane canlıdır. Connectome vb. projeler beyin sinir ağlarının araştırılmasına verilen önemi göstermektedir.

Yönetim bilimi: İşletme yönetiminde başarının anahtarının ağ bilimi olduğu yavaş yavaş anlaşılmaktadır. Şirket içi haberleşme ağlarının haritaları, şirket içi iletişim kopuklukları vb. sorunların kolaylıkla ortaya çıkartılmasına yardımcı olmaktadır. Örneğin, Wu ve arkadaşlarının ağ bilimi temelli araştırmaları, çalışanların verimlilikleriyle haritadaki konumlarının doğrudan ilintili olduğunu göstermiştir. Bu nedenle ağ bilimi yordamlarıyla şirketlerin gerçek ağ haritalarının çıkartılması konusunda hizmetler sunan onlarca danışmanlık firması ortaya çıkmıştır. IBM ve SAP gibi eskiden kurulmuş firmalar da danışmanlık ürünlerini ağ bilimi araçlarını içine alacak şekilde genişletmiştir. Bugün ağ biliminin etkin kullanımı işletmelerin sağlıklı bir şekilde yönetilebilmesi için olmazsa olmaz bir kriter olarak kabul edilmektedir. Bu bakımdan ağ biliminin sağladığı bilimsel veriler, müessesenin tüm ayrıntılarını ortaya koyup zamanında alınan doğru kararlarla şirket verimliliğinin artmasına yardımcı olan bir araç olarak bir nevi üst yönetimin elindeki mikroskop gibidir.