

# Büyük Veri Kahramanı **Veri Bilimci**

Son birkaç yıldır özellikle Batı Amerika'da iki kavram sıkça kullanılır oldu: "Büyük veri" ve "veri bilimi". 21. yüzyıl biliminin giderek veri ağırlıklı hale gelmesi nedeniyle, büyük ve karmaşık veri kümelerinden bilgi ortaya çıkarabilmek için yeni yeni analiz araçlarına ihtiyaç duyulmaya başlandı.

1997 yılında NASA araştırmacıları Michael Cox ve David Ellsworth tarafından ilk defa kullanılan "büyük veri" ifadesi, gittikçe kapasitesi artan verilerin kontrolünün ve yorumlanmasının oluşturacağı problemlere dikkat çekmek için kullanılmıştı.

Büyük veriyi büyük yapan önemli üç faktör vardır: İşlenen verinin hacmi, verinin karmaşıklığı ve çeşitliliği ve bilgi giriş çıkış hızı. Büyük veri, bilinen

ve çok kullanılan veri tabanı sistemlerinin işleyemeyeceği kadar geniş ve karmaşık veri kümelerinden oluştuğu için bu veriler istenilen düzeyde ne saklanabilir, ne de analiz edilebilir. Bu veri ancak doğru veri analiz araçları ve yetkin bir veri bilimciyle buluştuğunda hızlı ve doğru analiz edilerek, bir şirketin ya da bir projenin geleceği için ihtiyaç duyduğu çıkarımların elde edilmesini sağlamaya yardımcı olur.





Büyük veri Google, Facebook ve LinkedIn gibi arama ve sosyal paylaşım platformlarının yüz milyonlarca hatta milyarlarca kişinin bilgilerini toplayıp şirketlerinin gelişimi ve müşteri memnuniyeti amacıyla kullanması ve bu yöntemle kullanıcı sayılarını artırması nedeniyle bir anda popüler oldu. Öte yandan uzun yıllardır ABD dahil birçok ülkedeki büyük şirketler müşterilerinden topladıkları verileri şirketlerinin gelişimi için kullanmıyor, çöpe atıyorlardı. Ancak sosyal paylaşım platformlarının son zamanlardaki başarısının oluşturduğu örnek sayesinde bu verilerin kullanılmaya değer olduğu ortaya çıktı ve “büyük veri” devrimi oldu.

Aslında temelinde analiz olan her bilim dalı (örneğin uzay bilimleri, tıp, iklim bilim, fizik, istatistik) çok fazla parametreye içeren değişkenlerle çalıştığı için, ortaya çıkan çok büyük (terabyte) veri setlerini analiz etmek zorunda kalır. Bu açıdan bilim dünyası büyük veri ile aslında çok daha önce tanışmıştı. Fakat hızla gelişen teknolojiye paralel olarak artık hayatın her alanında daha büyük ve karmaşık

veri setleri ile başa çıkmak zorunda kalıyoruz. Sosyal medyada yer alan karmaşık veri setlerinin kullanım alanlarına şöyle bir örnek verilebilir. Facebook’un veri bilimcisi Carlos Diuk, iki insanın bir ilişkiye başlayıp başlamayacağını tahmin edebiliyor. Diuk’a göre “İlişkiye başlayacak iki kişi arasındaki Facebook iletilerinin sayısı ilişki başlamadan önceki 100 gün boyunca yavaş fakat düzenli bir artış gösteriyor. İlişki başladığında ise iletiler azalmaya başlıyor. Bir ilişki başlamadan 12 gün önce günde en yüksek 1,67 ileti gözlemlenirken, ilişki başladıktan sonraki ilk 85 günde ileti sayısının günde 1,53’e kadar düştüğü gözlemleniyor. Muhtemelen çiftler birlikte daha fazla zaman geçirmek istiyorlar ve dış dünyadaki etkileşimleri, sosyal medyadaki etkileşimlerine oranla daha da artıyor.”

Günümüzde yenilikçi bir fikri olan bir kişi rahatlıkla bir akıllı telefon uygulaması yazabilir ya da yazdırabilir. Yaklaşık 2 milyara yakın akıllı telefon kullanıcısı olduğunu düşünürsek, küçük bir reklam kampanyasıyla bir akıllı telefon uygulamasını belki 5000 kişinin telefonuna yüklemeye

si gayet olasıdır. Bundan sonra uygulamanın sahibi iki şey yapabilir: Bir, yeni müşterileri gelmesini bekleyebilir. İki, bir veri bilimci tutup programını kimler yüklemiş, bu kişiler ağırlıklı olarak hangi ülkelerde ve hangi şehirlerde, kaç yaşında, cinsiyetleri ne, programın kullanım sıklığı günün hangi saatinde artmış, programı kullanırken telefon yatay mı yoksa dikey mi tutulmuş gibi bilgileri kullanarak programını geliştirebilir ve kısa zamanda 50 bin ya da 500 bin kullanıcıya ulaşabilir.

Örneğin Google’ın bazı ülkelerdeki grip salgınını, salgın başlar başlamaz haber verebilmesi (<http://www.google.org/flutrends/>) yine veri biliminin kullanılması ile mümkün oluyor. İnsanlar grip olduğunda Google’da grip ile ilgili arama yapmaya başlıyor. Eğer bu arama sayıları belli bir bölgede hızla artıyorsa Google grip salgınının başladığını herkesten önce fark edebiliyor. Öte yandan aramalarda kullanılan sözcükler insanların ait oldukları kültürler, yaşayış tarzları gibi birçok konuda bilgiler verir ve hatta o kültürün geleceği hakkında dahi öngöründe bulunulmasını sağlayabilir.



## Kime Veri Bilimci Denir?

Aslında veri bilimi çok yeni bir alan olduğu için tanımı da kesin değil. Veri bilimci basitçe herhangi bir yazılım mühendisinden daha iyi istatistik ve herhangi bir istatistikçiden daha iyi yazılım mühendisliği bilen kişi olarak tanımlanabilir. Dolayısıyla yazılımcılar veri bilimi için ilk sırada tercih edilmiyor. Her ne kadar yazılımcılar çok iyi ve çok hızlı bilgisayar kodu yazıp yazılım üretebilseler de, veri bilimcinin esas görevi “çok iyi istatistik bilgisiyle verileri kullanarak sorunları çözmek hatta daha önceden sorun olduğu bile tahmin edilemeyen şeyleri öngörerek çözüm üretmek” olarak tanımlanabilir. Daha genel bir tanım yapmak gerekirse, veri bilimci kendini bilgisayar bilimi, istatistik, analitik modelleme ve matematik konularında geliştirmiş, çalıştığı alan ile ilgili kuvvetli sezgileri olan, güçlü iletişim yeteneğine sahip kişidir. Veri bilimci sadece sorunları fark etmekle kalmaz, aynı zamanda organizasyon için en önemli sorunları tespit etme konusunda da ayrı bir yeteneğe sahiptir.



Dr. Umut A. Yıldız

Selçuk Topal

Dr. Umut A. Yıldız, NASA Jet İtki Laboratuvarı'nda (JPL) astrofizik ve veri bilimi alanlarında araştırma yapıyor. Ankara Üniversitesi Astronomi ve Uzak Bilimleri Bölümü'nde lisansını, Groningen Üniversitesi'nde yüksek lisansını, Leiden Üniversitesi Gözlemevi'nde de yıldız oluşumu ve molekül astrofiziği alanında doktorasını tamamladı. Herschel Uzak Teleskobu'nu kullanarak yıldız oluşum bölgelerinde ilk kez su ve oksijen molekülleri keşfeden takımda yer aldı.

Selçuk Topal, lisans ve yüksek lisans öğrenimini Ankara Üniversitesi Astronomi ve Uzak Bilimleri Bölümü'nde

tamamladı. Milli Eğitim Bakanlığı Yurtdışı Doktora Bursu kazanarak 2010 yılında Oxford Üniversitesi Astrofizik Bölümü'nde doktora çalışmalarına başladı. Alanı gereği büyük veri setlerinin analizi ile yakından ilgilenen Topal, galaksimiz dışındaki diğer galaksilerde bulunan yıldız oluşum bölgelerindeki moleküler gazın fizikini ve kinematikini araştırdığı doktora çalışmalarını bitirmek üzere.

Selçuk Topal ve Dr. Umut Yıldız bir yılı aşkın bir süredir “Astronomi ve Bilimde Kariyer” başlıklı e-konferans dizileri düzenleyerek astronomiyi ve genel anlamda bilimi ülkemizin genç nesli arasında yaymayı amaçlayarak ortaokul ve liselere MEB Bilgi İşlem Dairesi'nin de desteğiyle seminerler veriyor. E-konferans taleplerini iletmek ve etkinliklerle ilgili bilgi almak için [www.ekonfastro.org](http://www.ekonfastro.org) adresi ziyaret edilebilir.



McKinsey Global Enstitüsü'nün 2011 yılında yazdığı bir rapor, önümüzdeki dört yıl içinde sadece ABD'de 140.000 ile 190.000 arasında veri analizi ve veri bilimci açığı olacağını ve bunun yanı sıra bu analizleri değerlendirecek yaklaşık 1,5 milyon yönetici kadrosu açığı olacağını öngörüyor. Bugün sadece iki yıllık geçmişi olan bir büyük veri analisti ABD'de yılda ortalama 200.000-300.000 dolar arasında kazanabiliyor. Bu demek oluyor ki, önümüzdeki yıllarda "veri bilimci" adıyla çok fazla iş pozisyonu açılacak ve yeterince yetişmiş insan olmayacak.

Peki henüz üniversitelerde veri bilimi bölümleri açılmamışken, bu insanlar nasıl yetişecek ve bu iş gücü açığı nasıl kapanacak? Bu büyük, karmaşık ve farklı formatlardaki veri setlerini analiz edip bilimsel olarak yorumlayabilecek nitelikte, disiplinler arası çalışabilecek insanlara ihtiyaç var. İşte bu nedenle zaten mesleği gereği bu tarz büyük veri setleri ile çalışan, astronomi, fizik, matematik gibi bilim dallarında uzmanlaşmış kişiler, büyük veri analizi için biçilmiş kaftan olarak görülüyor. 2008 yılında ABD'de yaşanan ekonomik kriz sonrasında üniversitelerdeki kadrolu araştırmacı sayısının düşmesi nedeniyle özellikle doktora yapmış fizikçiler ve astronomlar kadro beklemek yerine yavaş yavaş endüstride bu tarz işleri tercih etmeye başladı. Zaten aşına oldukla-

rı bir araştırma alanı olduğundan, başarıları kısa sürede Silikon Vadisi'nde duyuldu ve şu anda veri bilimci kadroları için en revaçta olan adaylar arasındalar. Yetkin bir veri bilimci olabilmek için matematik, algoritma, veri tabanları, istatistik ve programlama gibi konularda uzmanlaşmış olmak gerekiyor. Bazı üniversitelerin veri bilimi alanında verdiği online kursları (örneğin Coursera, Udacity) takip etmek ve şu an revaçta olan yazılımları ve programları (örneğin Python, Hadoop, R) öğrenmek büyük verilerin analizi konusunda kişisel gelişim için iyi bir başlangıç olacaktır.

İlerleyen yıllarda çok daha önem kazanacak olan bu yeni bilim ve meslek dalı, üniversitelerin özellikle sayısal alan mezunları için iyi bir alternatif olacak. Arzu eden öğrencilerin veri bilimi konusunda kendilerini geliştirmesini ve meslek seçenekleri arasına veri bilimci olmayı da eklemelerini tavsiye ediyoruz. Çünkü büyük veriye sahip olan ve bu veriyi doğru ve hızlı bir şekilde işleyip analiz eden, aynı zamanda geleceğin de sahibi olacak.

#### Kaynaklar

- Cox, M. & Ellsworth, D., "Application-Controlled Demand Paging for Out-of-Core Visualization", 1997.
- Mattmann, C., "A vision for data science", *Nature*, Sayı 493, s. 473-475, 2013.
- <http://www.businessinsider.com/facebook-knows-when-youre-going-to-fall-in-love-2014-2>
- <http://www.google.org/flutrends/>
- [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)
- <http://mollweide.files.wordpress.com/2013/05/big-data.jpg>