DATA SCİENCE

Veri Bilimi: 21. Yüzyılın En Seksi İşi

Veri odaklı dünyada, veri bilimcileri sıcak bir meta olarak ortaya çıktı. Veri biliminde en iyi yeteneği bulmak için kovalamaca devam ediyor. Zaten uzmanlar, veri bilimindeki milyonlarca işin, hazır yetenek eksikliği nedeniyle boş kalabileceğini tahmin ediyor. Nitelikli veri bilimcileri için küresel arama, yalnızca istatistikçiler veya bilgisayar bilimcileri için bir arama değildir. Aslında firmalar, konu uzmanlığına, yazılım programlama ve analitik konusunda biraz deneyime ve olağanüstü iletişim becerilerine sahip çok yönlü bireyler arıyorlar.

Dijital ayak izimiz son 10 yılda hızla genişledi. Dijital evrenin boyutu 1995'te kabaca 130 milyar gigabayttı. 2020'de bu sayı 40 trilyon gigabayta çıkacak. Şirketler, dijital dünyada gezinmek için ihtiyaç duyulan milyonlarca değilse de yüz binlerce yeni işçi için rekabet edecek. **Prestijli Harvard Business Review'un veri bilimini 21. yüzyılın en seksi işi olarak** adlandırmasına şaşmamalı .

McKinsey Global Institute tarafından hazırlanan bir rapor, veri ve analitik için büyük yetenek kıtlığı konusunda uyarıyor. 2018 yılına gelindiğinde, yalnızca Amerika Birleşik Devletleri, derin analitik becerilere sahip 140.000 ila 190.000 kişi ile etkili kararlar almak için büyük veri analizini kullanma bilgisine sahip 1,5 milyon yönetici ve analist sıkıntısıyla karşı karşıya kalabilir.

Dijital devrim hayatımızın her alanına dokunduğundan, davranışlarımızı öğrenmekten yararlanma fırsatı şimdi her zamankinden daha fazla. Doğru veriler göz önüne alındığında, pazarlamacılar alışkanlık oluşumumuza gizlice göz atabilirler. Nöroloji ve psikolojide yapılan araştırmalar, alışkanlıkların ve tercihlerin nasıl oluştuğunu ve Target gibi perakendecilerin bundan kâr elde etmek için dışarı çıktığını ortaya koyuyor. Ancak perakendeciler bunu ancak kendileri için çalışan veri bilimcileri varsa yapabilirler. Amazon.com'un eski baş bilim adamı Andreas Weigend, "Bu nedenle, günümüzde istatistikçileri işe almak bir silahlanma yarışı gibi" dedi.

C-suite yöneticilerini veri ve analitiklerin faydaları konusunda ikna etmeye hala ihtiyaç var. Görünüşe göre üst yönetim, analitik odaklı planlamanın potansiyeli hakkında bilgi sahibi olma konusunda orta yönetimin bir veya iki adım gerisinde olabilir. Wharton'da Müşteri Analitiği Girişimi'ni yöneten Profesör Peter Fader, yöneticilerin üst yönetime verilerle etkileşime girmek zorunda kalmadan ulaştığını biliyor. Gerçek değişimin, yöneticiler veri ve analitik konusunda bilgili olduklarında olacağına inanıyor.

Veri ve analitikte lider olan SAP, bir anketten, örneklemindeki yanıt veren firmaların %92'sinin veri varlıklarında önemli bir artış yaşadığını bildirdi. Aynı zamanda, dörtte üçü firmalarında yeni veri bilimi becerilerine olan ihtiyacı belirledi. Accenture, veri bilimcilerine olan talebin yalnızca 2015 yılında arzı 250.000 ile geçebileceğine inanıyor. 2014 yılında KPMG tarafından 150 yöneticiyle yapılan benzer bir anket, katılımcıların %85'inin verileri nasıl analiz edeceğini bilmediğini ortaya koydu. KPMG İngiltere dijital ve analitik başkanı Alwin Magimay, Mayıs 2015'teki bir röportajda, *çoğu kuruluş, veri ve analitiğin işlerini nasıl dönüştürebileceğini tam olarak anlamadıkları için noktaları birleştiremiyor , dedi.*

Forbes için yazan Bernard Marr, yetersiz analitik yeteneğiyle ilgili endişeleri de gündeme getiriyor. *Bu bilgiyi analiz etmek ve yorumlamak için gerekli becerilere sahip yeterli sayıda insan yok - onu ham sayısal (veya diğer) verilerden eyleme geçirilebilir içgörülere dönüştürmek - herhangi bir Büyük Veriye dayalı girişimin nihai amacı* . Bernard, Gartner'ın %50'den fazlasının veri biliminde kurum içi uzmanlık eksikliğini bildirdiği iş liderleriyle yaptığı bir anketten alıntı yapıyor.

Bernard, analitik ihtiyacı için kitle kaynak kullanımına yönelen Walmart hakkında bilgi verdi. Walmart, tescilli verilerini analiz etmek için bir yarışma düzenlemek için Kaggle'a yaklaştı. Perakendeci, kısa bir mağaza listesinden satış verileri sağladı ve rakiplerinden promosyon planlarına dayalı daha iyi satış tahminleri geliştirmelerini istedi.

Veri bilimci sıkıntısı göz önüne alındığında, işverenler yetenek için en yüksek dolarları ödemeye hazır. McKinsey'de müdür olan Michael Chui bunu çok iyi biliyor. Bir röportajda "Veri bilimi her şirketle alakalı hale geldi... Bu tür yetenekler için bir savaş var" dedi. Örneğin Paul Minton'ı ele alalım. Bir restoranda 20.000 dolarlık servis masası yapıyordu. Üniversitede matematik okudu. Bay Minton, her şeyi değiştiren üç aylık bir programlama kursu aldı. 2014 yılında San Francisco'da bir web girişimi için veri bilimcisi olarak 100.000 doların üzerinde para kazandı. *Hemen altı rakam... Bana göre şaşırtıcıydı,* dedi Bay Minton.

Bay Minton son derece şanslı olabilir mi, yoksa bu kadar yüksek maaşlar norm mu? Şansın bununla pek ilgisi yoktu; New York Times, bir yazılım mühendisinin ortalama temel maaşının 100.000 dolar ve veri bilimcileri için 112.000 dolar olduğunu bildirdi.

BÖLÜM 2

Birini Veri Bilimcisi Yapan Nedir?

Artık kitapta ne olduğunu bildiğinize göre, bazı tanımları yazmanın zamanı geldi. Her yerde kullanımlarına rağmen, fikir birliği Büyük veri ve Veri Bilimi kavramlarından kaçınır. Soru, Veri bilimcisi kimdir?oldukça canlı ve bazıları yalnızca kendi disiplinlerini veya akademik alanlarını korumakla ilgilenen bireyler tarafından tartışılıyor. Bu bölümde, bu tartışmaları ele almaya ve Büyük veri veya Veri biliminin dar bir şekilde yorumlanmış bir tanımının, yakın zamanda ortaya çıkan alana yönelen yüz binlerce kişinin dışlanmasına neden olacağını açıklamaya çalışıyorum.

Herkes bir veri bilimcisini sever,Simon Rogers (2012) Guardian'da yazdı. Bay Rogers ayrıca, sayı çarpmasına yönelik yeni keşfedilen sevginin izini, Google'dan Hal Varian'ın bir alıntısına kadar takip etti. Önümüzdeki on yıldaki seksi iş istatistikçiler olacak.

Hal Varian istatistikçileri seksi olarak adlandırırken, gerçekte kastettiği şeyin veri bilimcileri olduğuna inanılıyor. Bu birkaç önemli soruyu gündeme getiriyor:

* Veri bilimi nedir?
* İstatistiklerden farkı nedir?
* Birini veri bilimcisi yapan nedir?

Büyük veri zamanlarında, bu kadar basit bir soru, Veri bilimi nedir?birçok cevapla sonuçlanabilir. Bazı durumlarda, bu cevaplar hakkındaki görüş çeşitliliği düşmanlıkla sınırlanır.

Veri bilimcisini, uygun araçları kullanarak Büyük veya küçük verileri analiz ederek sorunlara çözüm bulan ve ardından bulgularını ilgili paydaşlara iletmek için hikayeler anlatan kişi olarak tanımlıyorum. Veri boyutunu kısıtlayıcı bir madde olarak kullanmıyorum. Belirli bir keyfi eşiğin altındaki bir veri, kişiyi daha az veri bilimcisi yapmaz. Veri bilimci tanımım, makine öğrenimi gibi belirli analitik araçlarla da sınırlı değil. Meraklı bir zihne, analitikte akıcılığa ve bulguları iletme yeteneğine sahip olduğu sürece, bu kişiyi bir veri bilimcisi olarak görüyorum.

Veri bilimini, veri bilimcilerinin yaptığı bir şey olarak tanımlıyorum. Yıllar önce, Toronto Üniversitesi'nde bir mühendislik öğrencisi olarak şu soruyla takılıp kaldım: Mühendislik nedir? Yüksek lisans tezimi konut fiyatlarını tahmin etme üzerine ve doktora tezimi de ev inşaatçılarının ne inşa edecekleri, ne zaman inşa edecekleri ve nerede yeni konut inşa edeceklerine ilişkin seçimlerini tahmin etme üzerine yazdım. İnşaat mühendisliği bölümünde, Diğerleri binalar, köprüler, tüneller tasarlamak ve şevlerin stabilitesi hakkında endişelenmek üzerinde çalışıyorlardı. Benim ve amirimin işi, sizin geleneksel bahçe çeşitliliği mühendisliğiniz değildi. Açıkçası, başkaları tarafından defalarca araştırmamın gerçekten mühendislik olup olmadığı soruldu.

Bu endişeleri doktora danışmanım Profesör Eric Miller ile paylaştığımda güldü. Dr Miller, bir ömür boyu kentsel arazi kullanımı ve ulaşımını araştırdı ve daha önce MIT'den doktora derecesi aldı. *"Mühendislik, mühendislerin işidir"*diye cevap verdi. Önümüzdeki 17 yıl boyunca, onun ifadesindeki bilgeliği anladım. Önce bir derece alarak ve ardından mühendislik mesleğini düzenleyen yerel meslek kuruluşuna kaydolarak mühendis olursunuz. Artık bir mühendissiniz. Tünel kazabilirsiniz; yazılım kodlarını yazın; bir iPhone veya süpersonik jetin tasarım bileşenleri. Sen bir mühendissin. Ve Dr Raghuram Rajan'ın yaptığı gibi, Uluslararası Para Fonu'nun (IMF) baş ekonomisti olarak bir finansal krize küresel müdahaleye liderlik ederken, bir mühendissiniz.

Profesör Raghuram Rajan, Hindistan Teknoloji Enstitüsü'nden elektrik mühendisliği alanında ilk derecesini aldı. Lisansüstü çalışmalarda ekonomi okudu, daha sonra prestijli bir üniversitede profesör oldu ve sonunda IMF'ye geldi. Halen Hindistan Rezerv Bankası 23. Başkanı olarak görev yapmaktadır. Entelektüel hünerinin yalnızca bir ekonomist olarak aldığı eğitimden kaynaklandığını ve bir mühendislik öğrencisi olarak öğrendiği temel bilgilerin onun problem çözme yeteneklerini geliştirmede hiçbir rolü olmadığını iddia edebilir mi?

Professor Rajan is an engineer. So are Xi Jinping, the President of the People's Republic of China, and Alexis Tsipras, the Greek Prime Minister who is forcing the world to rethink the fundamentals of global economics. They might not be designing new circuitry, distillation equipment, or bridges, but they are helping build better societies and economies and there can be no better definition of engineering and engineers—that is, individuals dedicated to building better economies and societies.

So briefly, I would argue that data science is what data scientists do.

Others have many different definitions. In September 2015, a co-panelist at a meetup organized by BigDataUniversity.com in Toronto confined data science to machine learning. There you have it. If you are not using the black boxes that makeup machine learning, as per some experts in the field, you are not a data scientist. Even if you were to discover the cure to a disease threatening the lives of millions, turf-protecting colleagues will exclude you from the data science club.

Dr Vincent Granville (2014), an author on data science, offers certain thresholds to meet to be a data scientist. On pages 8 and 9 in Developing Analytic talent, Dr Granville describes the new data science professor as a non-tenured instructor at a non-traditional university, who publishes research results in online blogs, does not waste time writing grants, works from home, and earns more money than the traditional tenured professors. Suffice it to say that the thriving academic community of data scientists might disagree with Dr Granville.

Dr Granville uses restrictions on data size and methods to define what data science is. He defines a data scientist as one who can easily process a So-million-row data set in a couple of hours, and who distrusts (statistical) models. He distinguishes data science from statistics. Yet he lists algebra, calculus, and training in probability and statistics as necessary background to understand data science (page 4).

Some believe that big data is merely about crossing a certain threshold on data size or the number of observations, or is about the use of a particular tool, such as Hadoop. Such arbitrary thresholds on data size are problematic because, with innovation, even regular computers and off-the-shelf software have begun to manipulate very large data sets. Stata, a commonly used software by data scientists and statisticians, announced that one could now process between 2 billion to 24.4 billion rows using its desktop solutions. If Hadoop is the password to the big data club, Stata's ability to process 24.4 billion rows, under certain limitations, has just gatecrashed that big data party.

It is important to realize that one who tries to set arbitrary thresholds to exclude others is likely to run into inconsistencies. The goal should be to define data science in a more exclusive, discipline- and platform-independent, size-free context where data-centric problem solving and the ability to weave strong narratives take center stage.

Given the controversy, I would rather consult others to see how they describe a data scientist. Why don't we again consult the Chief Data Scientist of the United States? Recall Dr Patil told the *Guardian* newspaper in 2012 that *a data scientist is that unique blend of skills that can both unlock the insights of data and tell a fantastic story via the data.* What is admirable about Dr Patil's definition is that it is inclusive of individuals of various academic backgrounds and training, and does not restrict the definition of a data scientist to a particular tool or subject it to a certain arbitrary minimum threshold of data size.

The other key ingredient for a successful data scientist is a behavioral trait: curiosity. A data scientist has to be one with a very curious mind, willing to spend significant time and effort to explore her hunches. In journalism, the editors call it having the nose for news. Not all reporters know where the news lies. Only those Who have the nose for news get the Story. Curiosity is equally important for data scientists as it is for journalists.

Rachel Schutt, News Corp'ta Baş Veri Bilimcisidir. Columbia Üniversitesi'nde veri bilimi dersi vermektedir. Aynı zamanda Doing Data Science adlı mükemmel bir kitabın da yazarıdır. New York Times ile yaptığı bir röportajda Dr Schutt, veri bilimciyi kısmen bilgisayar bilimcisi, kısmen yazılım mühendisi ve kısmen istatistikçi olarak tanımladı (Miller, 2013). Ancak bu, ortalama bir veri bilimcisinin tanımıdır. *"En iyiler"* , *"gerçekten meraklı insanlar, iyi sorular soran ve yapılandırılmamış durumlarla başa çıkmakta ve bu durumlarda yapı bulmaya çalışan düşünürler olma eğilimindedir."*

BÖLÜM 3

Veri Madenciliği Hedefleri Oluşturma

Veri madenciliğinde ilk adım, alıştırma için hedefler belirlemenizi gerektirir. Açıkçası, cevaplanması gereken kilit soruları belirlemelisiniz. Ancak, temel soruları belirlemenin ötesine geçmek, alıştırmanın maliyet ve faydalarıyla ilgili endişelerdir. Ayrıca, veri madenciliğinden elde edilen sonuçların beklenen doğruluk ve yararlılık düzeyini önceden belirlemelisiniz. Para sorun olmasaydı, gereken cevapları almak için gerektiği kadar para atabilirdiniz. Bununla birlikte, maliyet-fayda dengesi, veri madenciliği uygulamasının amaçlarını ve kapsamını belirlemede her zaman etkilidir. Sonuçlardan beklenen doğruluk düzeyi de maliyetleri etkiler. Veri madenciliğinden yüksek düzeyde doğruluk daha pahalıya mal olur ve bunun tersi de geçerlidir. Ayrıca, belirli bir doğruluk seviyesinin ötesinde, egzersizden fazla bir şey kazanmıyorsunuz, azalan getiriler göz önüne alındığında. Bu nedenle, istenen doğruluk düzeyi için maliyet-fayda dengesi, veri madenciliği hedefleri için önemli hususlardır.

Veri Seçme

Bir veri madenciliği uygulamasının çıktısı büyük ölçüde kullanılan verinin kalitesine bağlıdır. Bazen, veriler daha fazla işlem için hazırdır. Örneğin, perakendeciler genellikle büyük müşteri satın alma ve demografik veri tabanlarına sahiptir. Öte yandan, veriler veri madenciliği için hazır olmayabilir. Bu gibi durumlarda, diğer veri kaynaklarını belirlemelisiniz ve hatta anketler dahil yeni veri toplama girişimleri planlamalısınız. Veri türü, boyutu ve toplama sıklığı, veri madenciliği uygulamasının maliyeti üzerinde doğrudan bir etkiye sahiptir. Bu nedenle, veri madenciliği için ihtiyaç duyulan ve soruları makul maliyetlerle cevaplayabilecek doğru veri türünü belirlemek kritik öneme sahiptir.

Verileri Ön İşleme

Verileri ön işleme, veri madenciliğinde önemli bir adımdır. Genellikle ham veriler dağınıktır, hatalı veya alakasız veriler içerir. Ek olarak, ilgili verilerle bile bazen bilgiler eksiktir. Ön işleme aşamasında, verilerin alakasız özniteliklerini belirler ve bu tür öznitelikleri daha fazla değerlendirmeden çıkarırsınız. Aynı zamanda, veri setinin hatalı yönlerinin tespit edilmesi ve bu şekilde işaretlenmesi gerekmektedir. Örneğin, insan hatası, sütunlar arasında yanlışlıkla birleştirmeye veya bilgilerin yanlış ayrıştırılmasına neden olabilir. Veriler bütünlüğü sağlamak için kontrollere tabi olmalıdır. Son olarak, eksik verilerle başa çıkmak için resmi bir yöntem geliştirmeli ve verilerin rastgele mi yoksa sistematik olarak mı eksik olduğunu belirlemelisiniz.

Veriler rastgele eksik olsaydı, basit bir çözüm seti yeterli olurdu. Ancak, sistematik bir şekilde veriler eksik olduğunda, eksik verilerin sonuçlara etkisini belirlemeniz gerekir. Örneğin, büyük bir veri setindeki bireylerin belirli bir alt kümesi, gelirlerini açıklamayı reddetmiş olabilir. Girdi olarak bireyin gelirine dayanan bulgular, geliri bildirilmeyen bireylerin ayrıntılarını hariç tutacaktır. Bu, analizde sistematik önyargılara yol açacaktır. Bu nedenle, eksik veri içeren gözlemlerin veya değişkenlerin analizin tamamından veya bölümlerinden çıkarılıp çıkarılmayacağını önceden düşünmelisiniz.

Verileri Dönüştürme

İlgili veri öznitelikleri tutulduktan sonra, bir sonraki adım, verilerin saklanması gereken uygun formatı belirlemektir. Veri madenciliğinde önemli bir husus, fenomeni açıklamak için gereken özniteliklerin sayısını azaltmaktır. Bu, verilerin dönüştürülmesini gerektirebilir Temel Bileşen Analizi (bölümün ilerleyen kısımlarında gösterilmiş ve açıklanmıştır) gibi Veri azaltma algoritmaları, önemli bir bilgi kaybı olmaksızın özniteliklerin sayısını azaltabilir. Ek olarak, çalışılan fenomeni açıklamaya yardımcı olmak için değişkenlerin dönüştürülmesi gerekebilir. Örneğin, bir kişinin geliri veri setine ücret geliri olarak kaydedilebilir; kiralık mülkler gibi diğer kaynaklardan elde edilen gelir; hükümetten destek ödemeleri ve benzerleri.

Genellikle değişkenleri bir türden diğerine dönüştürmeniz gerekir. Sürekli gelir değişkenini, veri tabanındaki her kaydın düşük, orta ve yüksek gelirli birey olarak tanımlandığı kategorik bir değişkene dönüştürmek akıllıca olabilir. Bu, temel davranışlardaki doğrusal olmayanları yakalamaya yardımcı olabilir.

Veri depolama

Dönüştürülen veriler, veri madenciliği için elverişli hale getiren bir biçimde saklanmalıdır. Veriler, veri bilimcisine sınırsız ve anında okuma/yazma ayrıcalıkları veren bir biçimde saklanmalıdır. Veri madenciliği sırasında, orijinal veritabanına geri yazılan yeni değişkenler oluşturulur, bu nedenle veri depolama şeması, veritabanından verimli bir şekilde okumayı ve veritabanına yazmayı kolaylaştırmalıdır. Verileri güvenli tutan ve ayrıca veri madenciliği algoritmasının farklı sunuculara veya depolama ortamlarına dağılmış veri parçalarını gereksiz yere aramasını engelleyen sunucularda veya depolama ortamlarında veri depolamak da önemlidir. Veri güvenliği ve mahremiyeti, veri depolamak için en önemli husus olmalıdır.

Madencilik Verileri

Veriler uygun şekilde işlendikten, dönüştürüldükten ve saklandıktan sonra veri madenciliğine tabi tutulur. Bu adım, parametrik ve parametrik olmayan yöntemler dahil olmak üzere veri analizi yöntemlerini ve makine öğrenimi algoritmalarını kapsar. Veri madenciliği için iyi bir başlangıç ​​noktası veri görselleştirmedir. Veri madenciliği yazılımının gelişmiş grafikleme yeteneklerini kullanan verilerin çok boyutlu görünümleri, veri kümesinde saklı olan eğilimlerin ön anlayışının geliştirilmesinde çok yardımcı olur.

*Bu bölümün ilerleyen kısımlarında veri madenciliği algoritmaları ve yöntemleri detaylandırılmıştır.*

Madencilik Sonuçlarının Değerlendirilmesi

Veri madenciliğinden sonuçlar çıkarıldıktan sonra, sonuçların resmi bir değerlendirmesini yaparsınız. Resmi değerlendirme, algoritmaların veri üretmede ne kadar etkili ve verimli olduğunu görmek için modellerin gözlemlenen veriler üzerindeki tahmin yeteneklerini test etmeyi içerebilir. Bu, "örnek içi tahmin" olarak bilinir. Ek olarak, sonuçlar geri bildirim için kilit paydaşlarla paylaşılır ve daha sonra süreci iyileştirmek için daha sonraki veri madenciliği yinelemelerine dahil edilir.

Veri madenciliği ve sonuçların değerlendirilmesi, ana paydaşlardan alınan geri bildirimler ışığında üretilen sonuçların kalitesini iyileştirmek için analistlerin daha iyi ve geliştirilmiş algoritmalar kullanması için yinelemeli bir süreç haline gelir.

BÖLÜM 4

Neden Uzun Ebeveynlerin Daha Uzun Çocukları Olmaz?

Daha uzun ebeveynlerin genellikle ebeveynlerinden daha uzun olması gerekmeyen uzun boylu çocukları olduğunu fark etmişsinizdir ve bu iyi bir şeydir. Bu, uzun boylu ebeveynlerden doğan çocukların mutlaka diğerlerinden daha uzun olmadığı anlamına gelmez. Durum böyle olabilir, ancak mutlaka kendi "uzun" ebeveynlerinden daha uzun değildirler. Bunun neden iyi bir şey olduğunu düşünüyorum, basit bir zihinsel simülasyon gerektiriyor. Uzun boylu ebeveynlerden doğan birbirini takip eden her nesil, ebeveynlerinden daha uzun olsaydı, birkaç bin yıl içinde, insanlar kendi iyilikleri için daha büyük mobilyalara, arabalara ve uçaklara ihtiyaç duyacak şekilde rahatsız edici derecede uzun olacaklardı.

Sir Frances Galton 1886'da aynı soruyu inceledi ve bugün regresyon modelleri olarak bildiğimiz istatistiksel bir tekniğe ulaştı. Bu bölüm, istatistiksel analizin beygir gücü haline gelen regresyon modellerinin çalışmalarını araştırıyor. İster akademik ister profesyonel alanlarda olsun, hemen hemen tüm ampirik araştırma arayışlarında, regresyon modellerinin veya bunların varyantlarının kullanımı her yerde mevcuttur. Tıp biliminde, daha etkili ilaçlar geliştirmek, operasyon yöntemlerini iyileştirmek ve küçük ve büyük hastaneler için kaynakları optimize etmek için regresyon modelleri kullanılmaktadır. İş dünyasında, kamu ve özel sektör kuruluşlarının tüketici davranışlarını, firma üretkenliğini ve rekabet gücünü analiz etmede regresyon modelleri ön plandadır.

Yüksek lisans tezimle ilgili bir hikaye anlatarak regresyon modellerini tanıtmak istiyorum. Bu hikayenin regresyon modellerinin faydasını açıklamaya yardımcı olabileceğine inanıyorum.

Açık Sonuçlar Departmanı

1999 yılında, konut amaçlı gayrimenkuller için hedonik fiyat modelleri geliştirme üzerine yüksek lisans araştırmamı bitirdim. 500.000 gayrimenkul işlemini içeren projeyi tamamlamak üç yılımı aldı. Savunmaya hazırlanırken eşim cömertçe beni üniversiteye götürmeyi teklif etti. Biz yoldayken, "Söyle bana, araştırmanda ne buldun?" diye sordu. Sonunda son üç yıldır neler yaptığımı açıklamam istendiği için çok mutlu oldum. "Eh, konut fiyatlarının belirleyicilerini araştırıyordum. Büyük evlerin küçük evlerden daha çok sattığını gördüm" dedim, tezin taslağını elimde tutarken yüzümde muzaffer bir ifadeyle eşime söyledim.

Otoyol için rampaya yaklaşıyorduk. Cümlemi bitirir bitirmez eşim aniden arabayı omzuna çevirdi ve frene bastı. Araba dururken bana döndü ve şöyle dedi: "Bunu bulduğunuz için size yüksek lisans derecesi verdiklerine inanamıyorum. Size daha büyük evlerin küçük evlerden daha çok sattığını söyleyebilirdim."

O anda, bariz sonuçlar bölümünde ders veren bir profesör gibi hissettim. Konut fiyatları hakkında yaygın olarak bilinen şeyin bana yüksek itibarlı bir üniversiteden yüksek lisans derecesi kazandıracağı konusunda şok olduğu için onu nasıl suçlayabilirim?

Önümüzdeki on dakikayı ona araştırmamın inceliklerini açıklamak için ayırabilmek için karımdan arabayı sürmeye devam etmesini istedim. Bunun yerine bana beş dakika verdi, bunun onu bile gerektirmeyebileceğini düşündü. Beşe yerleştim ve bir sonraki dakikayı düşüncelerimi toplamakla geçirdim. Ona araştırmamın sadece konut fiyatları ile konut birimlerinin büyüklüğü arasındaki ilişkiyi bulmadığını, aynı zamanda bu ilişkilerin büyüklüğünü de keşfettiğimi açıkladım. Örneğin, bu bölümde daha sonra açıklayacağım bir terim olan diğer her şeyin eşit olduğunu buldum, ek bir tuvalet, ek bir yatak odasından daha fazla konut fiyatına katkıda bulunur. Aksi belirtilmedikçe, bir evin fiyatındaki marjinal artış, ilave bir tuvalet için ilave bir yatak odası için olduğundan daha yüksektir. Daha sonra Toronto'daki emlak komisyoncularının bu bulguyu gerçekten takdir ettiğini öğrendim. Ayrıca eşime metro gibi ulaşım altyapısına yakınlığın konut fiyatlarının artmasına neden olduğunu açıkladım. Örneğin, metroya daha yakın olan evler, uzaktakilerden daha fazla satıldı. Ancak, otoyollara veya otoyollara yakın evler diğerlerinden daha ucuza satıldı. Benzer şekilde, büyük alışveriş merkezlerine yakınlığın konut fiyatları üzerinde doğrusal olmayan bir etkisi olduğunu da keşfettim. Alışveriş merkezlerine çok yakın (2,5 km'den az) bulunan evler diğerlerinden daha ucuza satılıyor. Ancak, alışveriş merkezine daha yakın (5 km'den az, ancak 2,5 km'den fazla) bulunan evler, daha uzakta bulunanlardan daha fazla satıldı. Ayrıca Toronto'daki konut değerlerinin şehir merkezinden uzaklaştıkça düştüğünü gördüm. Ayrıca eşime metro gibi ulaşım altyapısına yakınlığın konut fiyatlarının artmasına neden olduğunu açıkladım. Örneğin, metroya daha yakın olan evler, uzaktakilerden daha fazla satıldı. Ancak, otoyollara veya otoyollara yakın evler diğerlerinden daha ucuza satıldı. Benzer şekilde, büyük alışveriş merkezlerine yakınlığın konut fiyatları üzerinde doğrusal olmayan bir etkisi olduğunu da keşfettim. Alışveriş merkezlerine çok yakın (2,5 km'den az) bulunan evler diğerlerinden daha ucuza satılıyor. Ancak, alışveriş merkezine daha yakın (5 km'den az, ancak 2,5 km'den fazla) bulunan evler, daha uzakta bulunanlardan daha fazla satıldı. Ayrıca Toronto'daki konut değerlerinin şehir merkezinden uzaklaştıkça düştüğünü gördüm. Ayrıca eşime metro gibi ulaşım altyapısına yakınlığın konut fiyatlarının artmasına neden olduğunu açıkladım. Örneğin, metroya daha yakın olan evler, uzaktakilerden daha fazla satıldı. Ancak, otoyollara veya otoyollara yakın evler diğerlerinden daha ucuza satıldı. Benzer şekilde, büyük alışveriş merkezlerine yakınlığın konut fiyatları üzerinde doğrusal olmayan bir etkisi olduğunu da keşfettim. Alışveriş merkezlerine çok yakın (2,5 km'den az) bulunan evler diğerlerinden daha ucuza satılıyor. Ancak, alışveriş merkezine daha yakın (5 km'den az, ancak 2,5 km'den fazla) bulunan evler, daha uzakta bulunanlardan daha fazla satıldı. Ayrıca Toronto'daki konut değerlerinin şehir merkezinden uzaklaştıkça düştüğünü gördüm. Örneğin, metroya daha yakın olan evler, uzaktakilerden daha fazla satıldı. Ancak, otoyollara veya otoyollara yakın evler diğerlerinden daha ucuza satıldı. Benzer şekilde, büyük alışveriş merkezlerine yakınlığın konut fiyatları üzerinde doğrusal olmayan bir etkisi olduğunu da keşfettim. Alışveriş merkezlerine çok yakın (2,5 km'den az) bulunan evler diğerlerinden daha ucuza satılıyor. Ancak, alışveriş merkezine daha yakın (5 km'den az, ancak 2,5 km'den fazla) bulunan evler, daha uzakta bulunanlardan daha fazla satıldı. Ayrıca Toronto'daki konut değerlerinin şehir merkezinden uzaklaştıkça düştüğünü gördüm. Örneğin, metroya daha yakın olan evler, uzaktakilerden daha fazla satıldı. Ancak, otoyollara veya otoyollara yakın evler diğerlerinden daha ucuza satıldı. Benzer şekilde, büyük alışveriş merkezlerine yakınlığın konut fiyatları üzerinde doğrusal olmayan bir etkisi olduğunu da keşfettim. Alışveriş merkezlerine çok yakın (2,5 km'den az) bulunan evler diğerlerinden daha ucuza satılıyor. Ancak, alışveriş merkezine daha yakın (5 km'den az, ancak 2,5 km'den fazla) bulunan evler, daha uzakta bulunanlardan daha fazla satıldı. Ayrıca Toronto'daki konut değerlerinin şehir merkezinden uzaklaştıkça düştüğünü gördüm. Ayrıca büyük alışveriş merkezlerine yakınlığın konut fiyatları üzerinde doğrusal olmayan bir etkisi olduğunu keşfettim. Alışveriş merkezlerine çok yakın (2,5 km'den az) bulunan evler diğerlerinden daha ucuza satılıyor. Ancak, alışveriş merkezine daha yakın (5 km'den az, ancak 2,5 km'den fazla) bulunan evler, daha uzakta bulunanlardan daha fazla satıldı. Ayrıca Toronto'daki konut değerlerinin şehir merkezinden uzaklaştıkça düştüğünü gördüm. Ayrıca büyük alışveriş merkezlerine yakınlığın konut fiyatları üzerinde doğrusal olmayan bir etkisi olduğunu keşfettim. Alışveriş merkezlerine çok yakın (2,5 km'den az) bulunan evler diğerlerinden daha ucuza satılıyor. Ancak, alışveriş merkezine daha yakın (5 km'den az, ancak 2,5 km'den fazla) bulunan evler, daha uzakta bulunanlardan daha fazla satıldı. Ayrıca Toronto'daki konut değerlerinin şehir merkezinden uzaklaştıkça düştüğünü gördüm.

Konut piyasaları araştırmasına katkılarımı açıklarken, karımın biraz etkilendiğini fark ettim. Ilık karşılamasının muhtemel nedeni, bulgularımın günlük deneyimlerimizden zaten bildiklerimizi doğrulamasıydı. Bununla birlikte, araştırmanın gerçek kattığı değer, bu ilişkilerin büyüklüğünü ölçmeye dayanıyordu.

Neden Gerileme?

Regresyon analizine bir sürü soru konabilir. Regresyon (hedonik) modellerinin ele alabileceği bazı soru örnekleri şunları içerir:

* Bir ev ek bir yatak odası için daha ne kadar satabilir?
* Parsel büyüklüğünün konut fiyatına etkisi nedir?
* Tuğla dış cepheli evler, taş dış cepheli evlerden daha mı ucuza satıyor?
* Bitmiş bir bodrum, bir konut biriminin fiyatına ne kadar katkıda bulunur?
* Yüksek voltajlı elektrik hatlarının yakınında bulunan evler diğerlerinden daha fazla mı yoksa daha az mı satıyor?