

Büyük veri analizi, görselleştirme ve risk yönetimi

Dünya dillerine, en son ve bir yenisi olarak “R” dili analitik dünyaya hükmederken biz hâlâ sıradan istatistik teknikleriyle mi devam edeceğiz?

Coşkun Küçüközmen

Sosyal medya, nesnelerin interneti ve büyük veri devrimleri sayesinde bilginin var olan gücünün daha da ön plana çıkması iş yapma tarzımızı radikal bir şekilde gözden geçirmemizi gerektiriyor. Büyük şirketlerin büyük bir kısmı piyasadaki paylarını korumak ve bir adım öne geçebilmek için fark yaratmak zorunda olduklarının bilincinde. En küçük bir bilgi kırıntısının bile ne kadar fark yaratacağının farkında olan birçok şirket bugün akademilerini çoktan kurmuş durumda.

Büyük şirketlerin bile kısa zamanda yok olabileceklerinin örnekleri giderek artıyor. Kodak şirketinin zaman içinde nasıl eriyip yok olduğu, Enron’un ve beraberinde Arthur Anderson isimli bağımsız denetim şirketinin batışı ya da Lehman Brothers isimli büyük yatırım bankasının kısa sürede nasıl iflas ettiği halen hafızalarda tazeliğini korumakta. Şirketlerin bugün teknoloji ile girdiklerini dönüşüm sürecini anlatan en güzel terim ise dijital dönüşüm terimi. Yani gelişen ve değişen iş ortamına ayak uydurmak için



gereken dönüşüm sürecinde teknolojiyi yoğun bir şekilde kullanmak. Bu süreçte veri toplamak, analiz etmek ve veriye dayalı karar vermek şüphesiz bu sürecin can damarı, odak noktası. Kâr marjlarının daraldığı, teknoloji sayesinde maliyetlerin ciddi şekilde düşürüldüğü, 3D yazıcılar ile üretim yerleşkelerinin kısa sürede mobil hale geldiği bir süreçte ayakta kalabilmenin yolu farkındalıktan geçiyor. Söz konusu farkındalığın ciddi bir maliyet gerektirdiği aşikâr.

Bu noktada çözüm nedir? Latince kökenli panacea kelimesi yazımız bağlamında her

Şirketlerin bugün teknoloji ile girdiklerini dönüşüm sürecini anlatan en güzel terim ise dijital dönüşüm terimi.

derde deva anlamına gelmektedir. İşte bu her derde deva olan çözüm şimdilik büyük veri ve bu verinin veri-madenciliği yoluyla görselleştirilmesi ve günlük hayatımıza girmesidir.

Bu bağlamda karşımıza bir kavram daha çıkmaktadır: Veri-Analitiği; yani büyük veriden anlamlı sonuçlar çıkarabilmek için gelişen teknolojilerle mevcut istatistikî yöntem ve araçların birleştirilerek şirketlerin, kurumların, bireylerin ihtiyaç duyabilecekleri bilgilerin elde edilebilmesi süreci. Son birkaç yıl içinde yayınlanan birçok çalışma, veri-bilimci, veri analitiği uzmanı, veri-uzmanı türün-

de mesleklerin önümüzdeki on yıllara damgasını vuracak cazip meslekler olduğunu vurguluyor.

Tekrar sorumuza dönelim, verinin geometrik seri hızında arttığı bir süreç işimize, yaşamımıza nasıl yansiyacak? Büyük veri olarak sınıflanan veriler iş yapma süreçleri içinde artık rutin olarak oluşuyor ve bizler de bunun bir parçası oluyoruz. İnternet kullanırken yaptığımız her tıklama, alışverişlerimiz, kredi kartı kullanımımız, cep telefonu ile yaptığımız görüşmeler, tedavilerimiz, banka hesaplarımız, yani hayatımızın neredeyse yarısından fazlası büyük veri kapsama alanına giriyor. Akla gelen ilk soru bu kadar büyük verinin nasıl depolandığı. Sosyal medyada da sıkça rastlanan bir fotoğrafta IBM firmasının geliştirdiği 5MB'lik süper kompüterin uçağa yüklendiği fotoğraftan yola çıkarak bu kapasitenin 14,000 katı büyüklükteki bir hafızanın bugün elimizdeki sıradan bir cep telefonuna sığdırıldığını düşünecek olursak sanırım bu konuda endişeye –en azından şimdilik– pek yer yok gibi görünüyor.

Veri ne kadar değerli?

Sorumuz şu: Veri nasıl değer yaratıyor ya da yaratmalı? Bugün birçok muhasebe ya da hukuk bürosu topladığı doküman vb. nedeniyle en fazla veriye sahip olan ve bunu kolayca gözlemleyebileceğimiz yerler arasında. Bu bağlamda artık tarihe karışmış olan bakkal veriye defterleri de bir veri kaynağı. Ancak bu kaynaklardan ham haliyle aradığımız bilgiye ulaşmak ve bunlardan anlamlı

sonuçlar çıkarmak neredeyse imkânsız. Neticede veriyi top- layandan ziyade veriyi ihtiyaca göre değerlendirebilenler bir katma değer yaratabiliyor. Ge- lişen teknolojinin imkânları ve fırsatları veri bilimlerini en aranan meslek ve insan grupları haline getirmiştir.

European Commission 2014 Temmuz ayında, büyük veri üzerine yeni bir strateji belgesi yayınlayarak (Towards a thriving data-driven economy, Brussels, 2.7.2014 COM(2014) 442 final) veri güdümlü bir ekonomiye geçişi destelemenin ve hızlandırmanın önemine işaret etti. Hatta çalışmanın başında “dijital veri, hesaplama ve otomasyon güdümlü yeni bir sanayi devrimine tanık oluyoruz” ifadesine yer verdi. Kanımızca veri güdümlü ekonominin, veri üzerine araştırma ve inovasyonu teşvik ederken, diğer yandan da yeni bir meslek olan veri analistlerinin iş dünyasına farklı bir ivme kazandıracığı kaçınılmaz görünüyor.

OECD tarafından 2014 yılında Interim Synthesis Report olarak yayınlanan ve Ekim 2015'te nihai halini alan Data-Driven Innovation for Growth başlıklı raporda ise büyük veri “ekonominin temel değeri olup, yeni endüstrileri, süreçleri, ürünleri teşvik edip, yeni rekabetçi avantajlardan bahisle büyük veriyi ekonomik gelişmenin merkezi konumundadır” ifadesine yer verilmiştir.

IDC (International Data Corporation) tarafından yapılan bir çalışmaya göre ise (Worldwide Big Data Technology and Services, 2012-2015 Forecast),

Veri güdümlü ekonominin, veri üzerine araştırma ve inovasyonu teşvik ederken, diğer yandan da yeni bir meslek olan veri analistlerinin iş dünyasına farklı bir ivme kazandıracığı kaçınılmaz görünüyor.

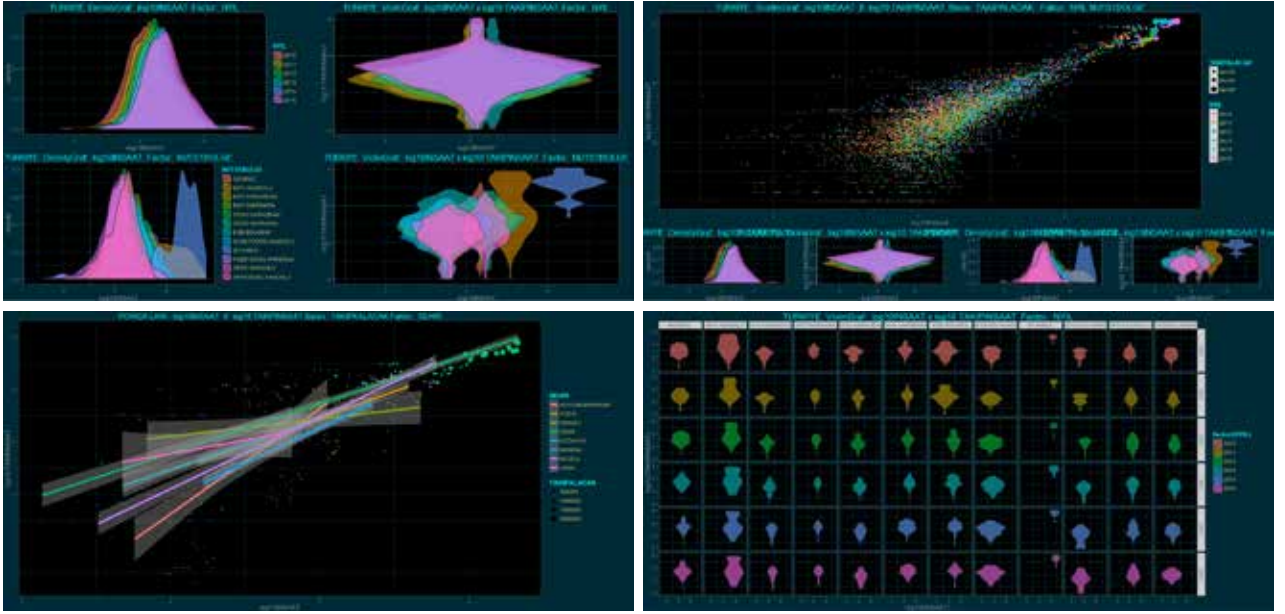
büyük veri teknolojisi ve hizmet pazarındaki büyümenin 2010 yılında 32 milyar dolardan, 2015 yılında 16.9 milyar dolara yükseldiği ifade edilmekte. Tüm bu gelişmeler ve analitik trendler 21 yy dünyasının en yeni ve en son elementin adını ortaya çıkarıyor: VERİ !

Veri madenciliği yoluyla görselleştirme

Bir şirket sunumunda olduğunuzu ve onlarca rakamın yer aldığı tabloları, art arda gelen grafikleri düşünün. Bir türlü her şeyi tek bir grafikte göremiyorsunuz. Hatta sizi ilgilendiren her şeyi tek bir grafik sayfasında görmek en büyük hayaliniz haline geliyor. İşte veri madenciliği yoluyla görselleştirme size bu imkânı sunuyor. Bu amaca uygun olarak geliştirilen yazılımların katkısı çözümleri grafik olarak gösterebilme yeteneklerinden kaynaklanıyor. Bu grafikleri yorumlamak ileri düzeyde analitik beceriler gerek-



Eylül 1956:
IBM ilk hard diskli bilgisayar olan 305 RAMAC'ı üretti.



tirmiyor. Konuyla ve sektörle ilgili ve makul seviyede profesyonel bilgisi olan herkes bu tür grafikleri üretme ve elde ettiği sonuçlardan analiz ve yorum yapabilme yeteneğine sahip oluyor. Böylece büyük veri veya herhangi bir veri setiyle ilgili sorun, uygun grafiklerin üretilebilmesine dönüşüyor. Bugünün yazılım teknolojisi bu tür grafiklemeyi ileri düzeyde sağlıyor. Bu yaklaşıma yoğunlaşanların risk yönetiminde bir değil, birkaç adım önde olacağı kesin! Aksi takdirde günümüz uygulamaları veriyi toplama, depolama ve iletişimi üzerine yoğunlaşmış. Bu veri kütesinden ne gibi yorumlar çıkartılabileceği yukarıda değinilen uzmanlık gereksinimleri yüzünden biraz gölgede kalmış. Yukarıda ifade edildiği üzere verinin yapılandırılması, analize elverişli hale getirilmesi ve elde edilen sonuçlardan işletmeye bir fayda sağlanması, bir değer elde edilmesi için en önemli kısmını oluşturmaktadır.

Tablolar tarihe karışmak üzere

Buraya kadar ifade edilenlerden de görüleceği üzere veri görselleştirmenin dijital dünyada bir sonraki büyük şey olacağını düşünmemiz için birçok haklı nedenimiz var. Karmaşık ve çok katmanlı zengin bilgiye bir bakışta ulaşılabilir ve görsel analiz doğru yapıldığında, gerçek zamanlı olarak sunulur ve güncellenebilir. Ayrıca, iletişimi de oldukça kolaydır. Ancak esas sorun karmaşık veri setlerinin görselleştirilmiş te olsa kimler tarafından nasıl yorumlanacağı. Yani bu tür bir yorum için ne tür bir becerinin gerektiği. Genel kabul görmüş yaklaşım veri analizi için çok üst düzeyde matematik, istatistik ve enformatik beceri ve deneyim gerektirdiği yönünde. Tabi ki tüm bu disiplinlerin tek bir kişide ya da yöneticide olması kolay değil. Bu nedenle yöneticiler ve veriler arasında bir ara yüz yani "Veri Analizi Uzmanlığı" gerekiyor. Söz konusu

Gerekli olan tek şey alışılan grafik tiplerinin dışındaki grafikleri (violin grafikler, radar grafikler, heatmap grafikler, matris grafikler gibi) yorumlama becerileri edinmektir.

veri analistlerinin temel işlevi veriler içinde saklı olan gizemli ilişkileri yöneticinin kolayca anlayabileceği ve yorumlayabileceği bir formata dönüştürmek olacak. Bunun için en etkin yöntemin verileri çok boyutlu grafikler halinde yöneticilerin anlayabileceği şekle dönüştürmektir. Bu yaklaşım üst düzey matematik, istatistik ve enformatik becerilere olan gereksinimi de ortadan kaldıracaktır. Gerekli olan tek şey alışılan grafik tiplerinin dışındaki grafikleri (violin grafikler, radar grafikler, heatmap grafikler, matris grafikler gibi) yorumlama becerileri edinmektir. Veri Analistleri, bu şekilde yaptıkları analizleri üst yönetim ile haberleşebilmek için ikna edici görselleştirmelere çevirmeli, gerekirse yeni iş metrikleri tanımlayarak performans ölçüm sistemlerini önerebilmelidir. Bunların yanında veri analistleri veriler ile yapılan iş arasında karar vericiye destek olacak önerileri sunabilecek iş-

letmecilik becerilerine de sahip olmalıdır.

Ücretsiz (licence-free) R yazılım görselleştirmenin merkezinde

Elinizde bulunan CIO Dergisi'nin Nisan 2015 sayısında Martin Heller'in "R ile büyük veriyi harmanlamak" başlıklı bir yazısı yayınlandı. Yazıda açık kaynak kodlu R programlama dili ile büyük veri üzerinde istatistikler ve grafiksel çıktılar alabilirsiniz deniliyor ve "R'nin gücü istatistiksel analizler gerçekleştirmek için aldatıcı biçimde basit gözükken çağrılarda yatıyor" ifadesine yer veriliyor. Açık kaynak kodlu ve ücretsiz (licence-free) R yazılımı Excel ile karşılaştırıldığında R hatırı sayılır oranda daha fazla istatistik ve grafik gücüne sahip. Bilhassa özel gereksinimler için paketler ekliyorsanız formüllerle ve rakamlarla dolu bir tabloya nazaran R script'ini doğrulamak çok daha kolay deniliyor. İşin ilginç yanı R kodlarının anlaşılır bir biçimde okunabilmesi. Diğer yandan R'nin devasa gücü ve mevcut R paketlerinin çokluğu epey göz korkutucu bir öğrenme sürecini ortaya çıkarabilir. Ancak R'yi öğrenirken ve kullanırken biraz istatistik bilgi ve birikime sahip olmanın çok faydası var. Diğer taraftan veri görselleştirmesinde R'nin bütün paketleri için bilgi sahibi olmak da gerekmiyor. Ggplot2 ve lattice gibi bazı ileri grafik paketlerini kullanmayı becermek yeterli olabiliyor.

Yukarıdaki alıntıyı niçin yaptık? Zira veriye dayanan görselleştirme süreçleri için mutlaka

bir model kullanmak gerekiyor. İşte bu modeli R yazılımı kullanarak oluşturmak mümkün. Burada dikkat edilmesi gereken husus analiz sonuçlarını doğru anlayıp, değerlendirebilen yöneticiler yetiştirmek. Ancak bu girişim kısa sürede sonuç verecek bir girişim değil. Üniversitelerin mutlaka Veri Analitiği alanına el atmaları gerekiyor. Dağınık bir yapıda da olsa bazı bölümlerden yapılacak ders ve hoca transferi ile lider bir akademisyen yönetiminde bu tür bölümlerin kurulması ve veri analistleri yetiştirmesi mümkün. Ülkemiz üniversitelerinin, buradaki hocaların ve yurtdışı bağlantılarının yeterli olduğu aşikâr, iş sadece orkestrasyona ve kolları sıvayarak işe başlama-ya kalmıştır.

Veri Görselleştirme, veri setlerindeki karmaşıklığın indirgenmesidir

İşletmelerin finansal fayda elde etmek ve pazar paylarını korumak için performanslarını çok iyi gözlemlmeleri ve gerekli kararları en iyi şekilde verebilmek için ölçümleri çok iyi yapabilmeleri gerekir. Daha da önemlisi veriye soracakları

Açık kaynak kodlu ve ücretsiz (licence-free) R yazılımı Excel ile karşılaştırıldığında R hatırı sayılır oranda daha fazla istatistik ve grafik gücüne sahip.

soruyu çok iyi bilmeleri gerekir. Performansı optimize edebilmeleri için ise bütünlük kuramsal uygulamalar ve entegre raporlamalarla "Object Tabanlı Entegre Kurumsal Performans Yönetimi (Object Oriented Complexity Management)" modeline geçmeleri gerekir. Bu model Kutlu Merih Fatma Çınar ve Coşkun Küçüközmen tarafından geliştirilen bir model olup detaylarına Google'dan Riskonometri ve Riskonomi anahtar kelimeleri sorgulanarak erişilebilir (ayrıca bkz. Çınar ve Küçüközmen, "Chaos, Complexity and Leadership 2013" içinde, 175-182, "Object Oriented Modelling of Corporate Complexity Performance Balance Card: CBBC", Springerlink, 2014).

Günümüz dünyasında "performansın zaman metriği değişmiştir, aynı zamanda performans düzeyi de artmıştır". Dolayısıyla gerçek zamanlı bir analizden söz ediliyorsa firmanın yarattığı gerçek değer ölçülmesine ve görselleştirilmesine olanak sağlayacak Grafik Data Mining tekniğine yoğunlaşmaları ve bunu öğrenmeleri gerekiyor. Bu bağlamda



Soldan Sağa: Coşkun KÜÇÜKÖZMEN Fatma ÇINAR ve Kutlu MERİH.

günümüz iş modelinin temel sorunu “hâlâ analitik dünyanın ölü diyagramlarına itibar ediliyor olmasıdır”. Yaşayan çok boyutlu işletmeleri kâğıt üzerindeki iki boyutlu ölü diyagramlara indirgemek faydadan çok zarara yol açmaktadır.

Kültür değişimi ve sonuç

Neticede çok kısa sürede değişik senaryoları hızlı bir şekilde test edip, doğru soruları sorabilen, gerekirse hızlı bir şekilde yön değiştirebilen veri kütesindeki aksiyonları şirketin performansına entegre edebilen yöneticilerin yetiştirilmesi mümkün olabilecektir. Böyle bir gelişimin gerek akademik gerek iş dünyasında ciddi bir iş görme kültürü değişimine neden olacağı da aşikâr. Bugün şirketlere ve CEO'lara profesyonel pencereden bakıldığında tepe yöneticilerinin (gençler dâhil) çoğunluğunun 20 yıl öncesinin yönetim anlayışı ve kültürünü halen sürdürmekte olduğu görülmekte. İşin ilginç yanı iyi

eğitim almış olmalarına rağmen bu gözlem konvansiyonel anlayıştan uzaklaşmalarının zor olduğunu göstermektedir. Çözüm bilgi toplumunu yönetebilme farkındalığından geçiyor. Uzun vadede bu yeni kültürün vazgeçilmez bir rekabet avantajı haline geleceğinin görülebilmesi. Gerekli yeteneklere, yetkin ve gelişime açık yöneticilere sahip olmayan organizasyonların yaşama ve başarılı olma şansı hızla azalacak. Diğer taraftan veri bilimcilerin işletme ve finans alanında yüksek lisans ve benzeri eğitimler yoluyla yöneticiliğe giden yolda diğer bilim dallarından gelenlere oranla daha fazla şansı olacağı da bir diğer gerçek olarak karşımıza çıkmakta.

İşte tüm bu sorulara yanıt verecek ve amaca hizmet edecek bir çalışmanın ürünü olan görseller bu sayfalarda bilgimize sunuluyor. Görsellere ilişkin daha kapsamlı çalışmalara @CORTEXIEN, @Riskonometri @Riskonomi @ExpertPortfolio twitter hesapları üzerinden

Bugün şirketlere ve CEO'lara profesyonel pencereden bakıldığında tepe yöneticilerinin (gençler dâhil) çoğunluğunun 20 yıl öncesinin yönetim anlayışı ve kültürünü halen sürdürmekte olduğu görülmekte.

ulaşmanız mümkün. Söz konusu görseller risk yönetimi sürecinin en önemli parametreleri olmaya aday. Zaten riski ölçme, daha doğrusu görselleştirme süreci risk yönetiminin en önemli girdisi değil mi?

Sibernetik dünyada hala konvansiyonel analitik alışkanlığımızdan kopamayıp, hala büyük veri ile zaman mı kaybedeceğiz? Dünya dillerine, en son ve bir yenisi olarak “R” dili analitik dünyaya hükmederken biz hâlâ sıradan istatistik teknikleriyle mi devam edeceğiz?

Yazımızı cevabını size bıraktığımız bir soruyla bitirelim. Grafik Data-Mining tekniği, sizce Kredi Risk Analizi ve Kur Savaşlarından zarar görmeden çıkmamıza ne kadar yardımcı olabilir?

(* Bu çalışmada büyük ölçüde aşağıdaki çalışmadan yararlanılmıştır:

C. Coşkun KÜÇÜKÖZMEN (2015), “Büyük veri, Şirketler ve Risk Yönetimi”, 24-46, ASOMEDYA, Eylül-Ekim 2015. 