



BÜYÜK VERİ VE VERİ GÖRSELLEŞTİRME

Sadullah ÇELİK*

Emrah AKDAMAR†

Öz

Gün geçtikçe bilgi üretimi üstel olarak artmakta ve dünyamız, insan bilgisini geliştirecek bir takım anlamlı bilgiler içermesi gerektiğine inandığımız verilerle dolup taşmaktadır. Bununla birlikte, anlamlı bilgi çıkarmak, büyük miktarda verinin düz metin veya geleneksel tabular biçiminde sunulması durumunda zor olmaktadır. Böylece, verilerin etkili grafik gösterimleri, insan görsel algılama yeteneklerini harekete geçirerek popülerlik kazanmaktadır. Veri görselleştirme, insan bilgisini güçlendirmek için soyut ve fiziksel olmayan verilerin bilgisayar tabanlı etkileşimli görsel temsillerinin kullanılmasıdır. Görselleştirmede, kullanıcıların beklenen etkinlikleri algılamaları ve keşfetmelerinin yanı sıra, veriler hakkında fikir sahibi olup beklenmedik yapıları görmelerine yardımcı olmak amaçlanır. Çok değişkenli veri görselleştirmesi için, görsel olarak analiz edilecek veri seti büyük boyutlardadır. Belirli bir bilgi görselleştirme türü olan çok değişkenli veri görselleştirme, pazarlama ve mühendislik tasarımından, endüstri ve finansal piyasalara kadar çeşitli alanlarda sayısız uygulamayı içermektedir. Bu nedenle görselleştirme birçok özellik arasındaki ilişkinin hayati öneme sahip olduğu aktif bir araştırma alanıdır. Bu çalışmada, öncelikle büyük veri konusu hakkında genel bilgiler verilmiştir. Daha sonra, yaygın olarak kullanılan görselleştirme teknikleri hakkında genel bilgiler verilerek, veri görselleştirmenin önemi üzerinde durulmuştur.

Anahtar Kelimeler: Büyük Veri, Veri Görselleştirme, Büyüklük, Çeşitlilik, Hız, Değer, Doğruluk.

BIG DATA AND DATA VISUALIZATION

Abstract

Knowledge production is exponentially increasing and our world is full of belief that we should have some meaningful information to improve human knowledge. However, making meaningful information is difficult when substantial amounts are presented in plain text or traditional tabular format. Thus, effective graphical representations of the data become popular by activating human visual perception capabilities. Data visualization is the use of computer based interactive visual representations of abstract and non-physical data to enhance human knowledge. Visualization is also intended to help users perceive and discover expected activities, as well as having an idea about the data and seeing unexpected events. For highly variable data visualization, the data set to be visually analyzed is large. Multivariate data, a specific type of information visualization, includes innumerable applications in a variety of areas ranging from visualization, marketing and engineering design to industry and financial markets. For this reason, visualization is an active area of research in which the relationship between many features is vital. In this study, firstly, general information about the big data topic is given. Later, general information about commonly used visualization techniques is given, and the importance of data visualization is emphasized.

Keywords: Big Data, Data Visualization, Volume, Variety, Velocity, Value, Veracity.

GİRİŞ

Endüstri Devrimini bundan 200 yıl önce, XX. yüzyılın son çeyreğine kadar yaşadık. Bu dönemde insan akli ve gelişen sofistike makineler dünyada büyük çaplı gelişmelerin

*Arş. Gör., Adnan Menderes Üniversitesi Nazilli İİBF Ekonometri Bölümü, sadullah.celik@adu.edu.tr

†Doktora Öğrencisi, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Bölümü, emrahakdamar1@gmail.com



yaşanmasını sağladı. Bu süreçte iş ve ticaret bağları güçlendi, mesafeler kısaldı, bilgi alışverişi, sermaye hareketleri ve pazar payları büyüdü, ülkeler ve toplumlar arası karşılıklı etkileşimde büyük artış yaşandı. XX. yüzyılın son çeyreğine gelindiğinde ise küresel bir fenomen haline gelen internet ile yeni bir devrime tanık olduk. İnternet devrimi, bilgi ve iletişim teknolojilerinin gücü, veri ağlarındaki bilgiye hızlı bir şekilde erişilebilmesini sağladı. Özellikle son 20 yılda internet iş dünyasında ve sosyal hayatta büyük değişimler yaşanmasına sebep oldu. Bugün geldiğimiz noktada, endüstri ve internet devrimleri bir araya gelerek teknoloji odaklı yeni bir dinamizmi ortaya çıkardı. General Electric (GE) bu yeni dönemi “Endüstriyel İnternet Çağı” ya da “Endüstrinin Geleceği” olarak adlandırmaktadır (geturkiyeblog.com/endustrinin-gelecegi, 19/09/2016). Çoğu kişi Endüstriyel İnternet Çağı’nı endüstrinin geçirdiği 3. büyük devrim olarak yorumlamaktadır. Endüstriyel İnternet, büyük ölçekli endüstriler ve işletmeler için birçok avantajı beraberinde getirmiştir. Büyük veri şu anda bile binlerce kişiye yeni iş imkanı sunmakta ve milyarlarca tasarruf sağlamaktadır. Fütüristik düşünürler kendilerine “robotlar insanları ele geçirecek” sorusunu sorarak, akıllanan makinelerin yeni özelliklere sahip bir iş gücüne ihtiyaç duyduğunu farketmişlerdir. Özellikle “Akıllı makinelerin küresel çapta bir ağa bağlanmaları, dünya çapında trilyonlarca sensörden gelen verinin internette buluşmasıyla ortaya büyük veri denilen bir okyanusu çıkarmıştır.” Fakat bu kadar büyük verinin niceliği başlı başına bir problem teşkil etmekte ve bu veriler arasında ilişki kurabilmek ve anlamlı sonuçlar elde etmek için çok fazla zaman ayırmak gerekmektedir. Son beş yıldır birçok şirket veri analizine olan insan ve varlık yatırımını neredeyse iki katına çıkarmıştır. Hiç şüphesiz, büyük veriyi anlamlı ve işe yarar bilgiye dönüştürmek için yenilikçi yazılımlara ve personel gücüne ihtiyaç vardır. Bugün artık büyük verinin analiz edilmesi, yorumlanması ve görselleştirilmesi için yeni iş alanlarının gelecekte giderek daha da büyüyeceği tahmin edilmektedir. 2020 yılında bağlı nesnelerin 50 milyarı bulması beklenmektedir. Bununla birlikte nesnelere akıllı da olsalar yine de kendilerinden daha akıllı olan insanlara ihtiyaç duyacaklardır (geturkiyeblog.com/endustriyel-internet-buyuk-veri-ve-operasyon, 09/09/2016).

1. LİTERATÜR TARAMASI

Büyük veri işletmelerden tüketicilere ve bilimden hükümete kadar hayatımızın bütün yönleriyle ilgili, devrim niteliğindeki bir konudur (Jagadish, 2014: 86-94). “Bilgi ekonomisinin petrolü olarak tanımlanan büyük verinin özellikle beşeri bilimlerde büyük bir dönüşüm başlattığı görüşü, son 10 yılın sıkça tekrarlanan bir söylemi haline gelmiştir” (Işıklı, 2014). Hükümetler ve sosyal bilimciler daha önce elde edemedikleri birçok bilgiye bugün büyük veri sayesinde ulaşmaya başlamıştır. Eğer analiz yapmak için doğru yöntem kullanılırsa büyük veri dünyayı algılama ve değiştirme şeklimizi değiştirip sorunların çözülmesinde büyük bir kolaylık sağlayacaktır. Bugün, başta Amerika olmak üzere birçok ülke en iyi uygulamaların geliştirilip belirlenmesi için büyük veriyi kullanmaya başlamıştır. Üniversitelerin araştırma programlarında analitik yöntemler üretilirken, bazı kuruluşlar da bunları teknolojiye adapte ederek hükümetlerin programlarında hayata geçirilmesini sağlamaktadır. Bunun en bariz örneğini 2010 yılında Amerika’nın Afganistan’daki savaşta kullandığını görmekteyiz. O dönemde henüz 28 yaşında olan Chris White, Harvard doktora sonrası okulunda büyük veri, istatistik ve makine öğrenimi üzerine çalışmaktaydı. White, Afganistan da devam eden savaşla ilgili bir brifing olarak, Afganistan’daki karanlık güçleri inceledi. Bu karanlık güçlerin eylemleri acımasız olmakla birlikte taktikleri ve bürokrasileri çok sofistikeydi. Bunlar öldürüyor, terör saçıyor, büyüyor ve kazanıyorlardı. ABD bunlara karşı koymak için büyük



veriden yararlanma fırsatı olduğunu duymuştu ve bu avantajı kullanmak istiyordu. White Nexus 7 olarak adlandırılan gizli bir DARPA (İleri Savunma Araştırma Projeleri Ajansı) programı çerçevesinde bir ekipte görevlendirilmişti. ABD ordusunun Afganistan'daki verilerin kaynağı; CIA, Ulusal Güvenlik Ajansı (NSA), GPS uyduları, cep telefonu kayıtları, cepheden gelen bilgiler, dijital finansal kayıtlar, güvenlik kamera kayıtları ve sosyal ağlardı. Bu bilgiler "devasa ve kıymeti bilinmeyen bir bilgi bütünü" oluşturuyordu. DARPA, White ve onun gibi bir grup araştırmacıyı bu verilerden anlamlı sonuçlar çıkarmaları için Afganistan'a yolladı. Grubun bazıları uydu verilerini ve karasal gözlem verilerini kullanarak trafik akışını (veya akmadığını ki bu da Taliban'ın ele geçirdiği bir kontrol noktasını veya yola döşenmiş bir bombayı gösteriyor) belirleyecekti. White ve ekibi Taliban ve El Kaide'nin finansının dijital izlerini sürmek, kölelik, seks, silah ve uyuşturucu ticaretinin kaynağını bulmak, bu işlemlerin nerede ve kimler tarafından gerçekleştiğini belirlemek için büyük veriyi kullanmıştır (popular science turkey, 03.11.2016:66-72). White'ın Memex adını verdiği arama motoru projesi sayesinde bir dedektifin iki haftada yapacağı araştırma çok kısa sürede yapılmakta ve bu sayede seks suçları trafiği saptanabilmektedir. Bugün Amerika da Manhattan Bölge Savcılığı insan ticareti ile ilgili tüm soruşturmalarda Memex'i kullanmış ve sadece 2016'nın ilk altı ayında 4.720 vakayı incelemiştir. Ayrıca Memex araçları sayesinde bugün IŞİD'in propagandası ve üye alım hareketlerini, naylon şirketlere para aklama arasındaki bağlantıları, yasadışı işgücü ya da silah kaçakçılığını, sosyal medyadaki sözcük ve fikirlerin haritalarını, bunların amacını harita üzerinde canlı olarak görmekte mümkün hale gelmiştir. Öyle ki; bugün artık hükümetler büyük veriyi kamu düzeni içerisinde etkin bir şekilde kullanmaya başlamışlardır. Yine de bilim adamları arasında bazı görüş farklılıkları bulunmaktadır. Bazı bilim adamları kamu politikalarının kanıta dayalı olduğunu söylerken bazıları da kötü kamu politikalarından dolayı bireylerin ihtiyaç duyduğu hizmeti alamadıklarını söyleyerek yeni yöntemlerin geliştirilmesini istemektedirler. Bugün bizlerden toplanan verilerin miktarı ve içerik mahremiyeti açısından da büyük bir tehlike yaratmaktadır. Çünkü bu verilerin yanlış ellerde kullanılma ihtimali bulunmaktadır. Bilim insanları bundan korunma yollarını uygulamalara geçirmek için seferber olmuştur. Hükümetler de bu veriye kimin erişebileceği konusunda halkı bilgilendirmeli ve güvenilir oldukları konusunda insanları ikna etmelidir (popsci.com/man-who-lit-dark-web, 01.10.2016).

Büyük veri bugün birçok alanda kullanılarak büyük başarılar imza atılmasını sağlamıştır. ABD Ulusal Okyanus Ve Atmosfer Teşkilatı (NOAA), ABD Ulusal Havacılık Ve Uzay Dairesi (NASA), sağlık sektöründeki birçok ilaç firması ve büyük ölçekli enerji firmaları bugün büyük miktardaki veriyi toplayarak bu veriden günlük bilgiler elde etmektedir. NOAA toplanan bu veriyi; iklim, ekosistem, hava durumu tahmini ve ticari amaç için kullanırken NASA daha çok havacılık araştırmaları için kullanmaktadır. İlaç şirketleri ve enerji şirketleri ise büyük veriyi ilaç deneyleri ve jeofizik analizleri yapmak için kullanmaktadır. Bugün New York Times büyük veri analizinde yeni nesil araçları kullanarak metin analizi ve web madenciliği yapmaktadır. The Walt Disney Company ise müşteri verilerini kullanarak müşteri alışveriş eğilimini tahmin etmektedir (Ohlhorst, 2013: 19-21). Yine MIT ve Birmingham Üniversin'deki araştırmacıların trafik verilerini kullanarak beş şehirde yaptığı bir çalışmaya göre az sayıda sürücünün bazı alternatif yolların tercih etmesiyle trafikteki yoğunluğun %30 oranında azaltılabileceğini bulmuşlardır. Bugün trafikte kaybedilen zaman sadece endüstriyel verimsizliğe değil bireysel verimsizliklere de neden olmaktadır. İstanbul da yapılan bir araştırmaya göre bir insan yılda yaklaşık olarak 125 saat trafikte durarak zaman kaybetmektedir. Bu da günde ortalama 8 saat



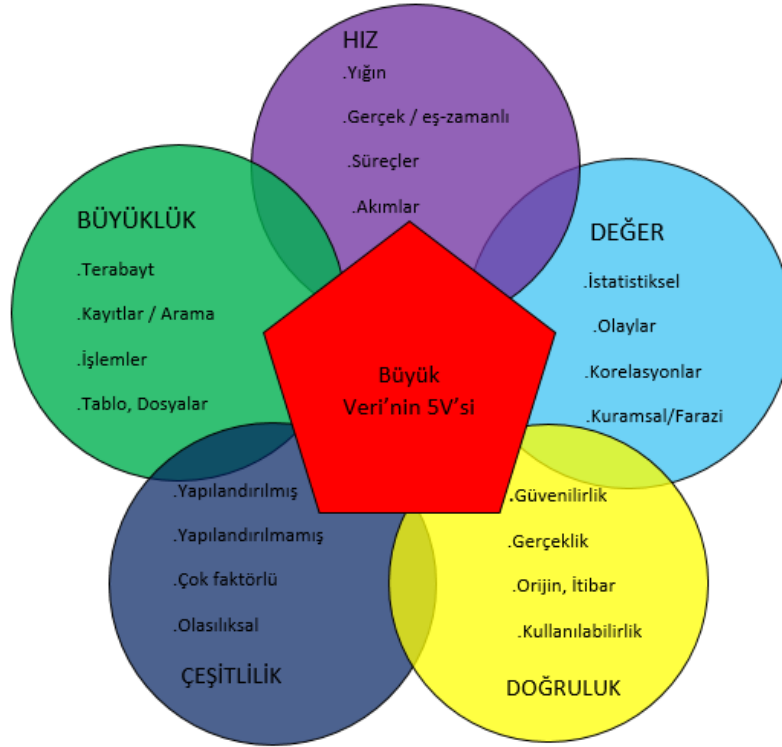
çalışan bir kişinin ortalama yılda 15 günlük mesaisini yolda geçirdiği anlamına gelmektedir (geturkiyeblog.com/buyuk-veri-trafigi-nasil-alt-eder, 20.05.2016).

2. BÜYÜK VERİ NEDİR?

Son on yılda, büyük veri giderek akademi, sanayi, hükümet ve diğer kuruluşların ilgisini çeken, küresel odaklı bir güç haline gelmiştir. "Büyük veri" terimi ilk olarak 1990'ların ortalarında bilimsel toplantılarda ortaya çıkarak 2008 yılında yavaş yavaş popüler hale gelmiş ve 2010 yılında tanınmaya başlanmıştır. Büyük veri yenilik, rekabet ve üretkenliğin baskın bir kaynağı olarak önerilmiştir (Manyika ve diğerleri, 2011). Daha geleneksel ölçme ve gözlem sistemlerine ek olarak birçok farklı sensör, mesajlaşma sistemi ve sosyal ağdan gelen büyük verilerin hızlı bir şekilde artması, daha şimdiden günlük hayatımızın büyük bir kısmını işgal etmektedir. IDC, Büyük Veri Teknolojilerini, yüksek hızda yakalama, keşif ve / veya analiz yaparak, çok geniş bir veri çeşidinden ekonomik olarak değer ayıklamak üzere tasarlanmış yeni nesil teknolojiler ve mimariler olarak tanımlamaktadır. Büyük verilerin üç temel özelliği vardır: verilerin kendisi, verilerin analitiği ve analiz sonuçlarının sunumu (Gantz ve diğerleri, 2012). Günümüzde veri büyüklüğü kentilyon bayta ulaşmıştır. Veri büyüklüğü her geçen gün katlanarak (üssel olarak) artmaktadır (en.wikipedia.org/wiki/Big_data, 10.10.2016). 2020'de internete bağlı cihazların 50 milyar olacağı tahmin edilmektedir. Bu kadar çok nesnenin internete bağlı olması Endüstriyel İnternet'in zorunluluğunu ortaya koymaktadır. Özellikle bulut tabanlı analizlerin yani "büyük veri" uygulamalarının endüstriyel alandaki adaptasyonları ile nesnelere, cihazlar ve üretim sistemleri çok daha yüksek verimde çalışabilmektedir (geturkiyeblog.com, 10.04.2017). Bugün internette yapılan her hareket veya tıklama veri oluşumuna neden olmaktadır. Bu hareketi gün içerisinde milyonlarca insanın yaptığını düşünürsek her saniyede devasa büyüklüğünde veri yığını oluşmaktadır. Fakat bunu sadece sosyal medya ile sınırlandırmak yanlış olur. Büyük veri kümeleri, sosyal ağlar, sensörler, dijital sinyaller, sağlık, finans, bilim ve benzeri her yerde toplanmaktadır. Başka bir deyişle, hemen hemen her veri kümesi büyük veri olma potansiyeline sahiptir. Çünkü birçok şirket için bilgi çok değerlidir. Bununla birlikte, geleneksel veritabanları ve analiz araçları kullanmadan yapılandırılmamış biçimdeki verileri, işleme ve analiz etmenin büyük zorlukları vardır. Bu nedenle, bu tür işlemleri gerçekleştirmek için özel sistemlere gereksinim duyulmaktadır. Bütün bu sistemlere genelde büyük veri denilmektedir. Büyük veri aslında yalnızca "büyük" değildir. Burada kastedilen aynı zamanda verinin çeşitli olmasıdır. Buradaki "büyük" kavramı yanıltıcı olup, "büyük" sözcüğü ile kastedilen, çok sayıda kaynaktan, eş zamanlı olarak gelen bol miktardaki çeşitli veridir. (geturkiyeblog.com/buyuk-veri-efsaneleri-2, 17.04.2017). "Büyük veri" kavramını daha iyi anlamak için onu oluşturan temel bileşenler üzerinde durmak gerekir. Büyük verinin 5V'si (Bkz. Şekil 1) olarak adlandırılan bu bileşenler: büyüklük (volume), çeşitlilik (variety), hız (velocity), değer (value) ve doğruluk (veracity)'dir.



Şekil 1: Büyük Veri'nin 5V'si



Kaynak: Zorlu, 2014.

Veri Büyüklüğü: Şirketler karar alma süreçlerinde çok büyük miktardaki veriyi kullanırlar. Verinin büyüklüğü üssel bir şekilde artmaya devam etmektedir. Bazı uzmanlar büyük verinin başlangıç noktasının petabayt olarak kabul ederken, birçok şirket ise bir terabayt ve bir petabayt arasındaki veri setlerini büyük veri olarak kabul etmektedir.

Çeşitlilik: Büyük veri yapısında fotoğraflar, tıklanma sayıları, e-postalar, sesler, videolar, HTML, PDF ve ekg verileri gibi çok çeşitli veri türlerini barındırmaktadır. Bu veriler yapılandırılmış (düzenlenmiş), yarı-yapılandırılmış ve yapılandırılmamış türdeki verilerdir (Bayrakçı, 2015). Büyük verinin büyük kısmı yapılandırılmamış yani klasik formatta satır ve sütuna yerleştirilmemiş verilerden oluşmaktadır.

Hız: Verinin üretilmesi, işlenmesi ve analiz edilmesi hızı devamlı artmaktadır. Daha yüksek hız ise üretilen verinin doğal gerçek-zamanında ve akan verinin iş süreçleri ile birleştirilmesi ihtiyacından kaynaklanmaktadır. Bugün, veri sürekli artan bir oranda üretilmektedir. Ancak bu veriyi geleneksel yöntemler kullanarak yakalamak, depolamak ve analiz etmek mümkün değildir. Özellikle çok kanallı anlık pazarlama gibi zamana duyarlı süreçler için, verinin iş değeri için eş zamanlı olarak analiz edilmesi büyük önem teşkil etmektedir.

Değer: Büyük veriyi açıklayan 3V (volume, velocity ve variety) tanımları yapıldıktan sonra verinin toplanıp işlenmesinden sonra işe yarar sonuçlar elde etmek için bu analizlerin bir değer” üretmesi gerekir. Dolayısıyla bu V'nin diğer 3V'nin (volume, velocity ve variety) birleşim noktası olduğu söylenebilir.

Doğruluk: Doğruluk belirli veri türleri ile ilgili güvenilirlik düzeyi anlamına gelir. Büyük veri için yüksek veri kalitesi önemli bir gereklilik ve mücadele arayışıdır. Fakat en önemli veri



temizleme yöntemleriyle bile bazı verilerin (hava durumu, ekonomi ya da bir müşterinin satın alma kararları gibi) doğasında var olan tahmin edilemezliği kaldırılamaz.

Bugün büyük veri ile ilgili birçok farklı tanım tapılmıştır. Bu tanımların farklı olmasının temel nedeni ise kullanıldığı alandan kaynaklanmaktadır. Büyük veri ile ilgili yapılan tanımların bazıları şöyledir:

McKinsey Global Institute (2011) raporuna göre ‘‘Büyük veri: veri setleri anlamına gelir ki bu veri setlerinin hacmi; yakalama, depolama, yönetme ve analiz etmek için geleneksel veri tabanı yazılım araçlarının yeteneğinin çok daha ötesindedir. ‘‘Büyük veri’’ terimini belli bir terabayt (binlerce gigabyt) sayısından daha büyük şekilde tanımlayamayız.’’

Teknoloji zamanla geliştikçe büyük veriyi ifade eden veri setlerinin boyutunun da artmasına neden olmuştur. Ayrıca büyük veri tanımı sektörün yaygın olarak bağlı bulunduğu yazılım araçlarına ve belli bir endüstride yaygın olan veri setlerinin boyutlarına bağlı olarak değişebilmektedir (Bkz. Tablo 1). Bunlara bağlı olarak bugün pek çok sektörde büyük veri birden fazla petabayt ile birkaç düzine terabayt (binlerce terabayt) aralığındadır (McKinsey, 2011). Örnek olarak Intel, büyük veriyi, ‘‘Bir haftada ortalama olarak 300 terabayt veri üreten firmalar için büyük veri fırsatlarının ortaya çıktığından söz ediyor’’ şeklinde tanımlamaktadır.

Tablo 1: Büyük Veri Nedir?

Sınıf	Büyüklik	Ne ile yönetilir?	Nerede saklanır?	Örnekler
Küçük	<10 GB	Excel, R	Bir makinenin belleği	Binlerce satış sayısı
Orta	10 GB-1 TB	Endekslenmiş dosyalar, monolitik veri tabanları	Bir makinenin diski	Milyonlarca web sayfası
Büyük	>1 TB	Hadoop, Spark, Dağıtık veri tabanları	Çok sayıda makinede saklanır	Milyarlarca web tıklanması

Kaynak: Narin, 2015.

3. GÖRSELLEŞTİRME (VISUALIZATION)

Veri görselleştirme, soyut bilgilerin analiz sonuçlarının grafikler, diyagramlar, tablolar, resimler veya animasyonlar oluşturularak sunulmasıdır. Görselleştirmede amaç, istatistiksel ve değişken bilgilerinin klasik formatta sunulan kompleks verileri, kolay algılanabilecek grafik arayüzler ile rahat anlaşılır hale getirmektir. İnsan algılama sistemi 3 boyut ile kısıtlanmakta ve daha yüksek boyutlu veri yapıları, insanın algılama sınırlarının dışına çıkmaktadır. Görselleştirme sayesinde çok boyutlu veriler boyut indirgemesi yapılarak 2 ya da 3 boyuta indirgenerek daha anlaşılır hale getirilir (Bilgin ve diğerleri, 2008). Yapılan araştırmalar insanların görsellere diğer herhangi bir uyarandan daha iyi tepki verdiğini göstermektedir. İnsan beyni görsel bilgiyi metinden 60.000 kat daha hızlı işler. Aslında, görsel veriler beyne iletilen bilgilerin yüzde 90'ını oluşturur (oracle.com/solutions/business-analytics/, 20.03.2017). Bugün, The New York Times gibi önde gelen medya kuruluşları, okuyucuları enfografi ile göz kamaştırmak ve metne gömülmelerini önlemek için haberlere dair bilgiler sağlamak amacıyla gelişmiş veri analizi ve yenilikçi bilgisayar grafikleri kullanmaktadır. ABD federal hükümeti ve endüstri kaynakları tarafından sağlanan veri setleri, bireylerin verilerin grafiksel olarak yorumlanmasını kendi başlarına bloglarda ve sosyal ağlarda yaratmalarına ve paylaşmalarına olanak tanımaktadır. Psikologlar ve beyin cerrahları, insanların grafik uyaranlara nasıl tepki



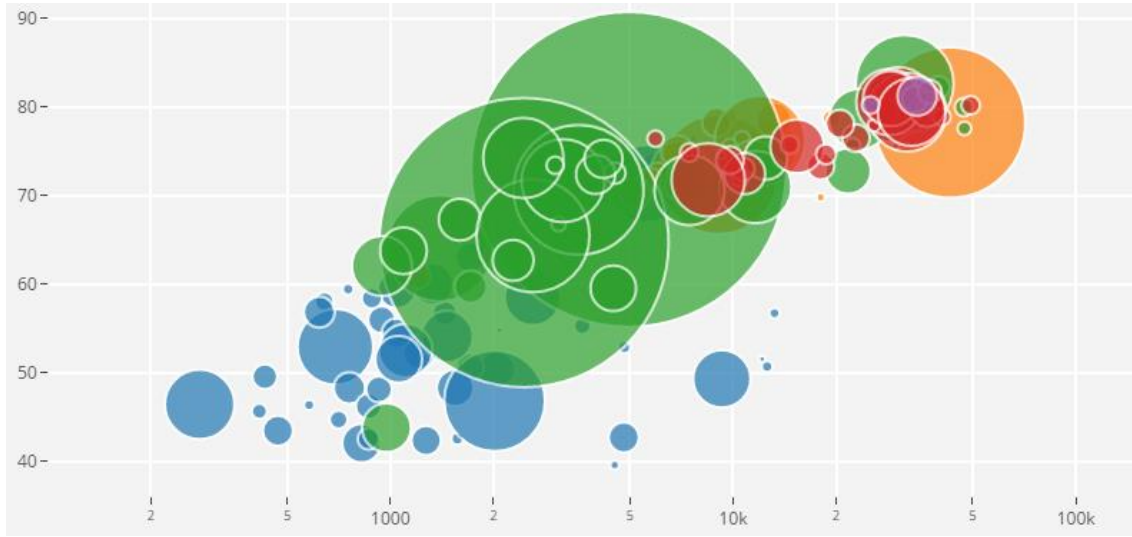
verdiklerini ve bilginin işlenmesinde önceki deneyimleri ortaya çıkarmak için kısa ve uzun süreli hafızayı nasıl kullandığını yoğun bir şekilde incelemişlerdir. Bu çalışmalar, doktorlar, pilotlar, finansal hizmet uzmanları, kolluk kuvvetleri ve askeri personel gibi her kesimden birer profesyonel kişiyi ele almış ve yapılan çalışmalar sonucunda görselleştirmenin algılamayı ve anlamayı büyük ölçüde kolaylaştırdığı görülmüştür. Bununla birlikte kararlar almak ve strateji geliştirmek için yeni bilgiler edinmek veri görselleştirmesi ile daha da önem kazanmıştır. Colin Ware; "Bilginin Görselleştirilmesi" adlı kitabında, "Görselleştirme ile tüm diğer duyulara kıyasla daha fazla bilgi edindik". "Beynin görsel bilgileri analiz etmek için kullandığı 20 milyar kadar nöron, bilişsel etkinliğimizin çoğunda temel bir bileşen olan bir model bulma mekanizması oluşturuyor" demektedir (Stodder, 2013).

Bugün yazılımcılar ve bilim insanları yeni görselleştirme araçları aramaktadır. Örneğin, Amazon, Twitter, Apple, Facebook ve Google gibi şirketler uygun işletme kararlarını vermek için veri görselleştirme yöntemini kullanmaktadır (Simon, 2014). Bu şirketler analitik raporlar ve veri görselleştirmeleri yapmak için İş Zekâsı sektörüne milyarlarca dolar ödemektedir. Şirketler İş Zekâsı sayesinde bir işe karar verme veya bir işlem yapmak için bilgi sağlayan herhangi bir türdeki veri görselleştirmeyi kullanmaktadır. İş Zekâsı'nın veri görselleştirmede önemli olmasının iki temel nedeni vardır: Birincisi, operasyonel ve ticari faaliyetler arasındaki ilişkileri ve gizli yapıları görselleştirmesidir. Bu sayede, kullanıcıların çalışma koşulları ve iş performansları arasındaki ilişkinin daha net görülmesine imkân sağlanır. Günümüzde rekabet gücü yüksek iş ortamlarında bu ilişkileri veriler arasında bulmak hiç olmadığı kadar önemlidir. İkincisi, veri görselleştirme iş ve işletim dinamiklerine çok yönlü bir bakış açısı sağlayarak müşteri memnuniyetinin ölçülmesinde belirgin bir şekilde kullanılmaktadır (Yuk ve diğerleri, 2017). Bugün veri görselleştirme de birçok teknik kullanılmaktadır. Bu tekniklerden en çok kullanılanlar şunlardır:

- *Hareket grafikleri (motion chart)*, büyük ve çok değişkenli verilerin etkili bir şekilde analiz edilmesine imkân tanır. Bu grafikler dinamik iki boyutlu kabarcık şemaları kullanarak etkileşime girer (Bkz. Şekil 2). Tasarlanan değişken haritalama nedeniyle lekeler (kabarcıklar - bu tekniğin merkezi nesnelere) kontrol edilebilir. Hareketli grafikler grafiksel veri araçları, Google, amCharts ve IBM Many Eyes tarafından sağlanmaktadır (Olshannikova ve diğerleri, 2015).
- *Sözcük bulut*, metin içerisinde geçen kelimelerin kullanım sıklığına (frekansına) bağlı olarak bunları şematik bir tabloda bir araya getirip frekansı en çok olan sözcükleri daha büyük gösteren bir tekniktir (Bkz. Şekil 3.). Bu teknik sayesinde pek çok sitenin, tezin, makalenin ve romanın vurguladığı anlamlar kolayca analiz edilebilmektedir (Sönmez, 2009). Aşağıdaki Şekil 3'te Tableau programı kullanılarak metin görselleştirmesi yapılmıştır.
- *Clustergram*, Hiyerarşik küme analizinde, kümelerin nasıl oluşturulduğunu görselleştirmek için dendrogram grafikleri kullanılır. Küme sayısı arttıkça, küme elemanlarının kümelerine nasıl atandığını incelemek için "clustergram" adlı alternatif bir grafik önerilmektedir (Bkz. Şekil 4). Bu grafik, k-araçları gibi hiyerarşik olmayan kümeleme algoritmaları için keşif analizinde ve gözlem sayısı dendrogramları pratik yapacak kadar büyük olduğunda hiyerarşik küme algoritmaları için kullanışlıdır (McKinsey, 2011).

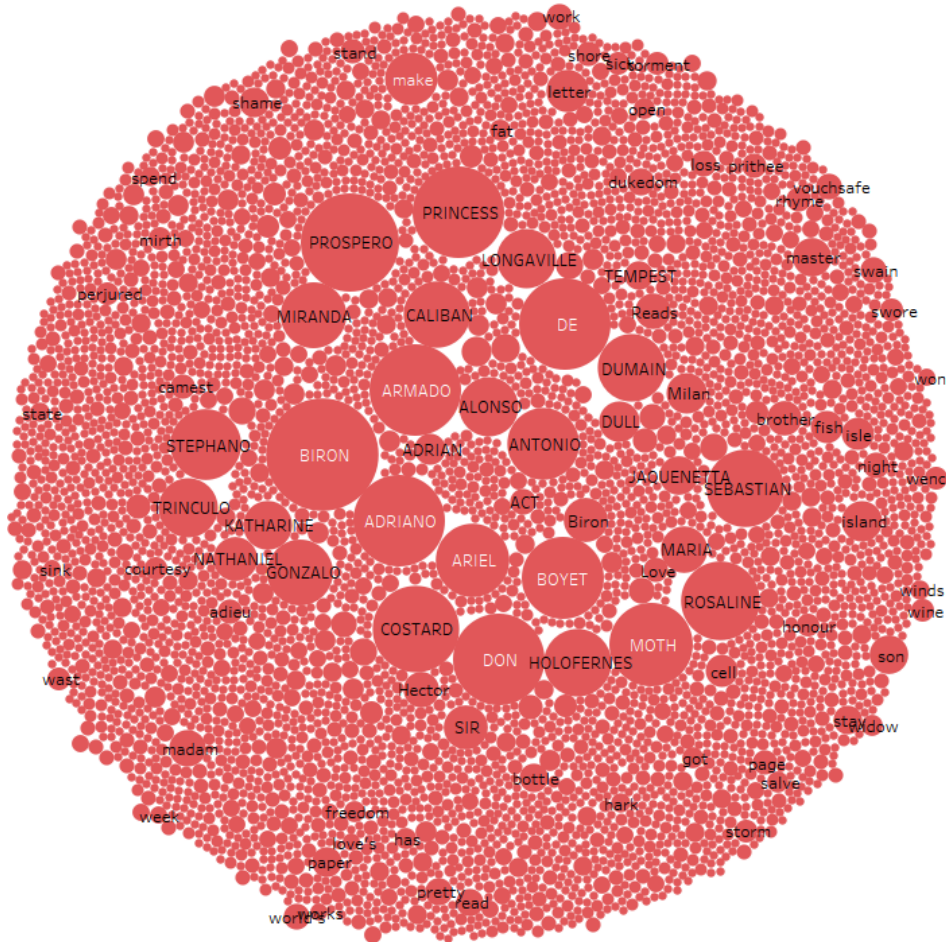


Şekil 2: Hareket Grafiği



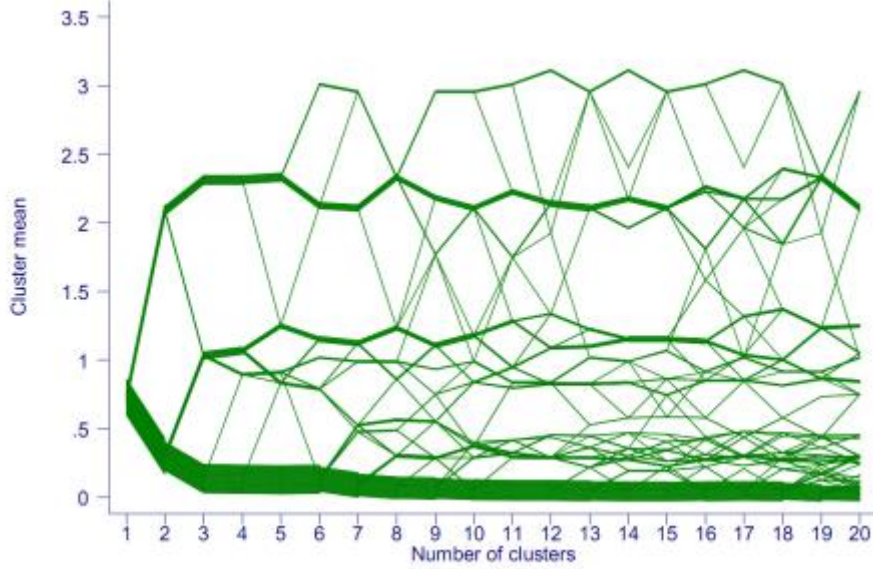
Kaynak: plot.ly/python/bubble-charts, 10.02.2017.

Şekil 3. Sözcük Bulutu (Metin Analizinde Kelime Görselleştirmesi)





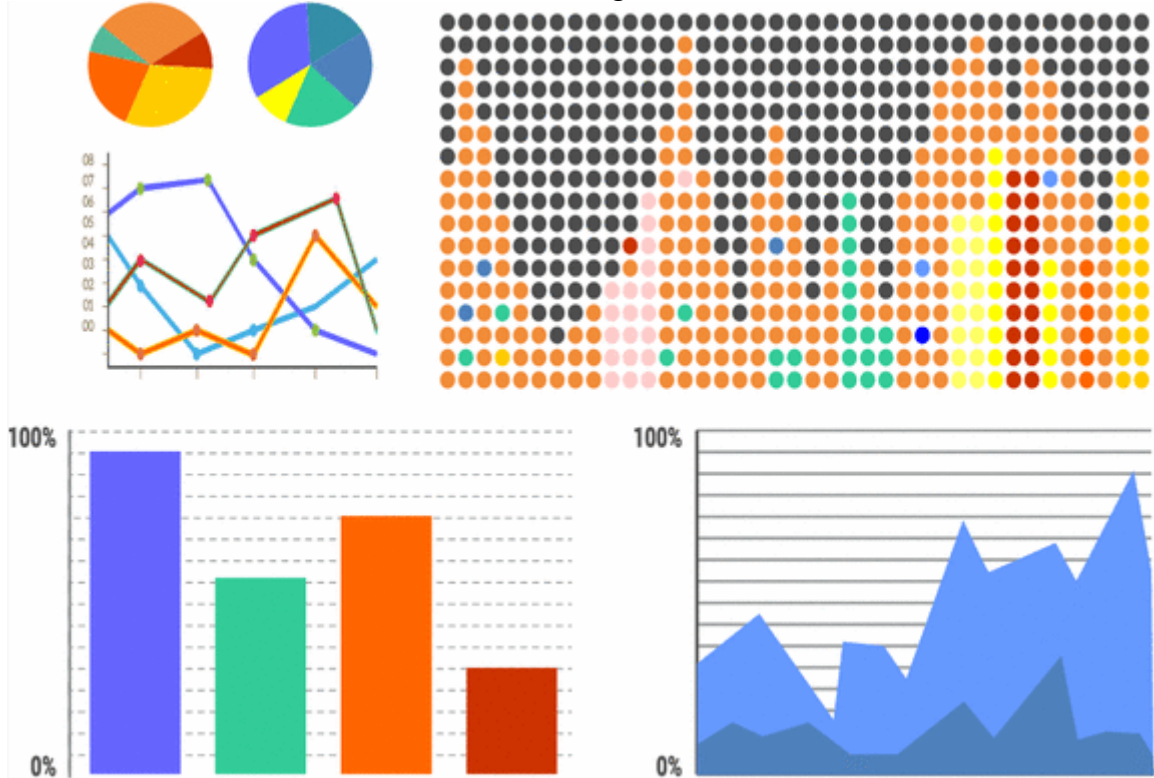
Şekil 4: Clustergram



Kaynak: McKinsey, 2011.

- Gösterge tablosu (Dashboard), bir veya daha fazla hedefe ulaşmak için gereken en önemli bilgilerin görsel bir gösterimidir. Gösterge tabloları konsolide edilmiş ve tek bir ekranda düzenlenmiştir (Bkz. Şekil 5). Bu sayede bilgiler tek bir bakışta izlenebilmektedir (Sandy Chiang, 2011).

Şekil 5: Gösterge Tablosu



Kaynak: Olshannikova ve diğerleri, 2015.



Bugün görselleştirme için kullanılmakta olan modern yöntem ve araçların tümü temel bilişsel psikoloji ilkelerine dayanmakta ve boyut, renk ve görsel nesnelere arasındaki bağlantıların manipüle edilmesi gibi temel kriterler kullanılmaktadır. İnsan bilişsel psikolojisi açısından Gestalt İlkeleri büyük önem taşımaktadır. Gestalt İlkeleri psikolojinin görsel algıya yönelik bir çalışma alanı olduğunu söylemektedir (Guberman, 2015:25-44). İnsanların, dünyayı, kurucu parçalardan ziyade bütünsel düzenlenmiş bir konfigürasyonda algıladıkları ileri sürülmektedir (Örneğin kişinin ilk önce ormanı algılaması ve daha sonra tekli ağaçların tümünü bir bütün olarak algılaması mümkündür). Ayrıca insan zihni boşlukları doldurur, belirsizliği önlemeye çalışır ve benzerlikleri ve farklılıkları kolayca tanır. Yakınlık yasası (Bir grup oluşturan nesnelere toplanması), benzerlik yasası (Nesneler birbirine benzerse algısal olarak gruplandırılır), simetri (Nesneleri simetrik şekiller olarak algılanması eğilimi), kapanış (Beynimiz tamamlanmamış objeleri kapatma eğilimindedir) ve şekil-zemin yasası (Görsel objelerin belirgin ve durgun rolleri) büyük verilerin görselleştirilmesinde dikkate alınmalıdır. Bu amaçla, en etkili görselleştirme yöntemi, çoklu ölçütleri en iyi şekilde kullanan yöntemdir. Aksi takdirde, çok fazla renk, şekil ve bağlantı, verilerin anlaşılmasında zorluklara neden olabilir veya bazı görsel öğelerin tanımda karmaşıklığa sebep olabilir (Olshannikova ve diğerleri, 2015).

TARTIŞMA VE SONUÇ

Büyük veri ve veri analizi ulaşım, elektrik, sağlık, havacılık, tarım, finans ve perakende gibi birçok alanda büyük değişimlerin yaşanmasına sebep oldu. Bugün şirketler, bankalar ve kamu kuruluşları her gün çeşitli kaynaklardan gelen: finans, mobil, sağlık, işlem, müşteri araştırma ve sosyal medya verisi gibi birçok veriyi veritabanlarında kolayca işleyebilmektedir. Bilişim ve bulut teknolojilerin de yaşanan gelişmeler ve çevrimiçi veri kaynaklarına anında erişilebilmesi yeni nesil güçlü araçların ortaya çıkmasına sebep oldu. Bu gücün analitik araçlarla birleştirilmesiyle akademisyenlere ve işletmecilere daha önce erişilemeyen veya kullanılmayan büyük veri kümelerine erişim ve kullanım imkânı sağladı. Bugün işletmeler ve hükümetler; metin analizi, makine öğrenimi, tahmini analiz, veri madenciliği, istatistik, doğal dil işleme ve görselleştirme gibi teknikleri kullanarak, daha önce kullanılmamış büyük miktardaki mevcut kurumsal verileri analizler ederek yeni anlayışlar kazanarak daha iyi ve hızlı kararlar alabilmektedirler. Bu çalışmada öncelikle büyük veri konusu hakkında genel bilgiler verilmiş ve daha sonra, yaygın olarak kullanılan görselleştirme teknikleri hakkında genel bilgiler verilerek, görselleştirmenin kuruluşlar ve devletler için önemi anlatılmıştır.

Veri görselleştirme, bilgisayar bilimlerinde oldukça yeni ve umut verici bir alandır. Desenleri, eğilimleri ve veri kümelerindeki ilişkileri ortaya çıkarmak için bilgisayar grafik efektlerini kullanır. Bugün gelişen teknoloji sayesinde büyük ve karmaşık veri setleri kolayca analiz edebilecek hale gelmiştir. Daha önceleri büyük ve karmaşık veri kümeleriyle karşılaşıldığında örnekleme yöntemine başvururken bugün artık verilerin hepsini (anakütleyi) basit bir şekilde işleme koyabilmekteyiz. Özellikle 19. yüzyıldan itibaren büyük ve karmaşık verilerle karşılaşıldığında örnekleme yöntemine başvurulmaktaydı. Fakat, örnekleme yöntemi bilginin az olduğu ve yüksek kapasitedeki dijital teknolojilerin yaygınlaşmadığı bir dönemde genellikle doğal karşılanıyordu. Bugün geliştirilen yeni nesil programlar bize verilerin tamamını kullanma imkânını sağlayarak daha önce sınırlı miktarda veri ile göremeyeceğimiz birçok ayrıntıyı



görmemize imkan sağlamıştır. Bu sayede örneklemin erişemediği alt kategorilerin ve altyapıların çok daha net bir görüntüsünü görmemize de yardımcıdır.

Büyük veri öncülük ettiği teknolojik dönüşümler sayesinde özellikle gelişmekte olan ülkeler için büyük fırsatlar yaratmaya devam etmektedir. Bugün birçok sektörlerdeki uygulamalarında da görüldüğü gibi, büyük veri, doğru yorumlandığında kurumlara ve toplumlara yeni fırsatların yaratılmasını sağlamaktadır. Büyük veri, iklim değişikliği, hastalık gözetimi ve doğal afet olaylarının, altınada yatan temel nedenleri saptamaya imkan sağlayarak ulaşım, sağlık ve pazarlama gibi birçok alanda toplumsal yaşamı kolaylaştırıcı bir etkiye sahip olmuştur. Diğer taraftan, büyük verilerin topluma sağladığı faydalar genellikle veri gizliliği, gizlilik ve güvenlik gibi konularda sınırlıdır.

Bugün teknolojideki gelişmelere rağmen büyük veri çağı henüz evriminin başlangıç aşamasındadır. Bu nedenle, büyük veri işleme yöntemleri büyük veri sorunlarını çözmek için gelişmekte ve yeni çözümler sürekli geliştirilmektedir. Bu ifadeyle, büyük veri dünyasının, karmaşık yapıları ve aralarındaki bağlantıların daha iyi anlaşılmasına yol açan çok disiplinli yöntemlerin ve tekniklerin geliştirilmesi gerektiği ifade edilebilir.

KAYNAKÇA

- Gantz, J., Reinsel, D. (2012), “The Digital Universe in 2020: Big Data”, Bigger Digital Shadows, and Biggest Growth in the Far East., Study report, IDC, December 2012. URL
- Guberman, S. (2015), “Guberman S. On Gestalt theory principles. GESTALT THEORY”, 2015;37(1):25–44.
- Işıklı, Ş. (2014), “Büyük Veri, Epistemoloji ve Etik Tartışmalar”, Online Academic Journal of Information Technology, Fall – Vol: 5 --/Num:17.
- Jagadishj, H.V., Gehrck, J., Labrinidis, A., Papakontantinou, Y. Patel, J., Rama-Krisnan, J.M.R. ve Shahabi, C. (2014), “Communications of the ACM.”, 57 (7): 86-94.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011), “Big Data: The Next Frontier for Innovation”, Competition, and Productivity.
- Ohlhorst. F. (2013), “Big data analytics: turning big data into big Money”, New Jersey, s.19-21
- Olshannikova, E., Ometov, A., Koucheryavy, Y., Olsson, T. (2015), “Visualizing Big Data with augmented and virtual reality: challenges and research agenda”, Journal of Big Data, 2:22. Doi:10.1186/s40537-015-0031-2.
- Simon, P. (2014), “The visual organization: data visualization”, Big Data, and the quest for better decisions. John Wiley & Sons.
- Sönmez, S. (2009), “Etiket Bulutu Analizi”, 25.11.2009.
- Stodder, D. (2013), “Data Visualization And Discovery For Better Business Decisions”, Third Quarter 2013.
- Tugay Bilgin, T., Yılmaz Çamurcu, A. (2008), “Çok Boyutlu Veri Görselleştirme Teknikleri”, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale.



Yuk, M., Diamon, S. (2017), “Data Visualization Why Business Intelligence is Important”.

İnternet Kaynakları

Bayrakçı, S. (2015), “Büyük Veri Nedir?”, <https://serkanbayrakci.wordpress.com/tag/big-data/>, (Erişim tarihi: 19.04.2015)

Narin, B. (2015), “Büyük Veri”, <http://es.slideshare.net/BilgeNarin1/big-data-24-mart-2015> (Erişim tarihi: 18.06.2017)

Sandy Chiang, A. (2011), “What is a Dashboard?”, <http://www.dashboardinsight.com/articles/digital-dashboards/fundamentals/what-is-a-dashboard.aspx/> (Erişim tarihi: 20.02.2017)

Zorlu, E. (2014), “Big Data”, <https://emrezorlu.com/2014/07/12/big-data/>, (Erişim tarihi: 25.07.2017)

<http://geturkiyeblog.com/endustrinin-gelecegi-ve-endustriyel-internet-devrimi/> (Erişim tarihi: 19.09.2016)

<http://geturkiyeblog.com/endustriyel-internet-buyuk-veri-ve-operasyon-optimizasyonu/> (Erişim tarihi: 19.09.2016).

<http://www.pressreader.com/turkey/popular-science-turkey/20161001/s.66-72>, (Erişim tarihi: 26.02.2017)

<http://www.popsci.com/man-who-lit-dark-web/> (Erişim tarihi: 01.10.2016)

<https://geturkiyeblog.com/buyuk-veri-trafigi-nasil-alt-eder/> (Erişim tarihi: 20.05.2016)

http://en.wikipedia.org/wiki/Big_data#Critiques_of_the_Big_Data_paradigm (Erişim tarihi: 10.10.2016)

<https://geturkiyeblog.com/dijital-sanayi-endustriyel-internet/> (Erişim tarihi: 10.04.2017)

<https://geturkiyeblog.com/buyuk-veri-efsaneleri-2/> (Erişim tarihi: 17.04.2017)

<https://www.oracle.com/solutions/business-analytics/data-visualization.html/> (Erişim tarihi: 20.03.2017)

<https://plot.ly/python/bubble-charts/> (Erişim tarihi: 10.02.2017)