



Sosyal Bilimlerde Büyük Veri Kullanımı, Veri Toplamada Akademik Çalışmalara Ne Tür Kolaylıklar Sağlayabilir?

What Kind of Facilities Can Provide the Use of Big Data in Social Sciences for Academic Studies in Data Collection?

Dr. Zübeyir ÖZÇELİK¹ - Doç. Dr. Ebru AYKAN²

Başvuru Tarihi: 29.08.2019

Kabul Tarihi: 09.09.2020

Makale Türü: Derleme

Öz

Bilgi işlem teknolojilerinin gelişimi ile birlikte zaman içerisinde veri toplama yöntemlerinde de değişiklikler gerçekleşmektedir. Günümüzde sosyal bilimlerde araştırmacıların veri toplama amacıyla sıklıkla başvurduğu anket, yüz yüze görüşme, gözlem yapma, deney gibi geleneksel veri toplama teknikleri, zaman içerisinde büyük veri kullanımına doğru evrilecektir. Bu çalışmada sosyal bilimlerde araştırmacıların; geleneksel yöntemlerle veri toplama teknikleri ile büyük veriden elde edilebilecek veriler arasında ne tür farklılıklarla karşılaşabilecekleri ele alınmıştır. Büyük veri ile veri elde ediminde nelere dikkat edilmesi gerektiği, büyük verinin veriyi üretenlere ve faydalananlara sunduğu avantajlar ve dezavantajlar değerlendirilmiştir.

Anahtar Kelimeler: *Büyük Veri, Akademik Çalışmalarda Büyük Veri Kullanımı, Veri Toplama Yöntemleri*

Abstract

With the development of IT technologies, there have been changes in data collection methods over time as well. It is clear that traditional data collection methods such as surveys, face-to-face interviews, observation and experiment, which researchers apply frequently in social sciences, will evolve towards the use of big data in time. In this study, what kind of differences the researchers in social sciences may encounter between the data that can be obtained from traditional data collection methods and big data, what are needed to pay attention to while acquiring data from big data; advantages and disadvantages provided by big data to the producers and beneficiaries of data are discussed.

Keywords: *Big Data, Big Data Use in Academic Studies, Data Collection Methods*

¹ Türkiye İstatistik Kurumu Başkanlığı, Ankara, zubeyirozcelik@tuik.gov.tr, ORCID: 0000-0001-7027-7396

² Kayseri Üniversitesi İzzet Bayraktar Uygulamalı Bilimler Yüksek Okulu, aykane@erciyes.edu.tr, ORCID: 0000-0003-3537-5235

Giriş

Günümüz bilgi işlem teknolojisinin hızlı gelişimi son 2 yılda üretilen verinin 2 yıl öncesinden geçmişe kadar üretilen tüm verilerin hacimleri ile eşit olduğu bir ortamda (Sağiroğlu, Sinanç, 2013, s. 42) araştırmacılar daha az emek ve daha az zaman harçayarak büyük veri kullanımı ile doğruya daha yakın sonuçlara ulaşabilme şansını elde edebileceklerdir. Francis X. Diebold, “*Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting*” adlı çalışmasında büyük verinin; fizik, biyoloji ve sosyal bilimler dâhil olmak üzere, birçok bilim alanında karşılaşılabilecek ve kullanıcılara yarar sağlayabilecek bir “fenomen” olarak bahsetmektedir (Diebold, 2013, s. 115). Hiç şüphesiz araştırma sonuçlarını şekillendirecek büyük verinin sağladığı kolaylıklar yanında elde edilmesinde ve kullanımında zaman içerisinde ortaya çıkan ve gelecekte de oluşacak belirli sınırlar bulunmaktadır. Bu çalışmada sosyal bilimlerde geleneksel veri toplama yöntemleri ile büyük veri kullanımı arasındaki farklılıkların araştırılarak değerlendirilmesi, araştırmacıların araştırmaya başlamadan önce araştırmanın tasarım ve yöntemini belirlerken geleneksel yöntemlerin haricinde büyük veri kullanımının ne tür fırsatlar ve sınırlılıkları olacağına değerlendirilmesine ışık tutmak amaçlanmaktadır.

Büyük Verinin Tanımı

Büyük veri; tüketicilerden işletmelere, bilim insanlarından devlet yönetimlerine kadar geniş bir yelpazeyi kaplayan (Jagdish vd., 2014) ve gündemde önemli yer tutan bir konudur. Vinod (2013, s. 96) büyük veriyi, verinin büyüklük olarak terabit veya petabitin yüzlerce katı olmasını tanımlayan bir olgu olarak ifade etmiştir. Rubistein (2013, s. 74) ise büyük veriyi; işletme, devlet ve organizasyonların dijital farklı veri setlerini bütünleştirerek istatistik ve veri madenciliği teknikleriyle gizli kalmış bilgilerin kullanılarak anlamlı sonuçlar elde edilmesi olarak tanımlamıştır. Büyük veri, geleneksel veri tabanı işleme yöntemlerinin uygulanmasıyla mümkün olmayan, farklı hacimlerdeki farklı veriyi tanımlayan yeni bir kavramı içermektedir (Gahi, Guennoun, Mouftah, 2016, s. 953). Büyük veri kavramının daha iyi anlaşılabilmesi için Gahi ve arkadaşlarının (2016) yapısal veri, yarı yapısal veri ve yapısal olmayan veri kavramları aşağıda açıklanmaktadır;

Yapısal veri: Girdileri, saklanması, işlenmesi, sorgulanması ve raporlaştırılması kolay olan veri türlerini açıklamaktadır. Yapısal veri ilişkisel veri tabanlarında belirli formatlarda tutulan ve kolayca ulaşılabilen ve belirli standartlar dahilinde dosyalanan ve depolanan veri türünü ifade etmektedir. Yapısal veri sürekli bir veriyi ifade eder ve analizi kolaydır.

Yarı yapısal veri: Yapısal özelliklerinin yanı sıra etiketleri, işaretleri ve farklı modelleri bünyesinde bulundurmaktadır. XML, HTML uzantılı metinler ve JSON programlama dilleri örnekler arasında yer almaktadır.

Yapısal olmayan veri: Önceden belirlenmiş bir formatı olmayan kitaplar, makaleler, belgeler, e-postalar, gps sinyalleri, resimler, banka hesapları, sosyal medya, idari kayıtlar, resmi kayıtlar gibi değişik formatlardaki metinlerden ve resim, ses ve video gibi medya dosyalarından ve depolanan kayıt türlerinden oluşmaktadır. Yapısal olmayan verilerde SQL yerine veri işleme süreçlerinde NoSQL (Not only SQL) gibi yeni sorgulama türleri kullanılmaktadır.

Veri işlenebildiği ve bilgiye dönüştürülebildiği sürece bir anlam ifade etmektedir. Günümüzde mevcut teknolojilerin kullanımı ile birlikte mobil cihazlar, yazılım kayıtları, kameralar, mikrofonlar, sosyal medya, internetteki tüm hareketler yapısal olmayan veriler kapsamında bilgi akışında işlenmek üzere depolanabilmektedir.

Büyük veri aşağıda sunulan 5V ile ifade edilen bileşenlerinin büyüklüğü ile açıklanmaktadır (Turan, 2016, s. 4);

1-) Hacim (volume): Makineler, akıllı okuyucular, uydular, cep telefonları, bilgisayarlar, video ve görüntüler, mailler, yazışmalar vb. kaynakların hepsi veri üretmektedir. Üretilen verinin hacmi gün geçtikçe artmaktadır. 2015 yılında küresel olarak 7,9 zetabayt olan veri hacminin 2020 yılında 38 zetabayt olacağı tahmin edilmektedir (Özdoğan, 2016, s. 14). Günümüzde veri akış hızı ve hacmi ile doğru orantılı olarak gigabayt, terabayt, petabayt, ekzabayt ve nihayetinde zetabaytlar düzeyinde oluşan verinin hacmi gündeme gelmiştir. (Sağıroğlu, 2017).

2-) Hız (velocity): Veri geçmiş zamana göre çok daha hızlı üretilmektedir. Günümüzde duran statik verilerden akışkan ve gerçek zamanlı dinamik veri üretim sürecine geçilmiştir. İnternete bağlanabilen ve gerçekleştirdiği faaliyetleri kayıt altına alabilen birbirlerine kablolu ya da kablosuz bağlanan ve birbirleriyle haberleşen cihazların birçoğu veri üretebilmektedir (Doğan, Aslantekin, 2016, s. 26). Örneğin; Bir boeing 737 uçağının motoru uçuş esnasında her 30 dakikada yaklaşık 10 terabayt veri üretmektedir (Big Data). Benzer şekilde her ay Youtube'a 4 milyardan fazla video yüklenmekte ve izlenmektedir. Twitter ortamına ise günde 200 milyondan fazla aktif kullanıcı 400 milyondan fazla tweet göndermektedir. Facebook'da ise günlük 30 milyardan fazla içerik paylaşılmaktadır (Özdoğan, 2016, s. 13). Dolayısıyla yukarıdaki örneklerde ifade edildiği düzeyde hızla akan veri, gerçek zamanlı (real time) işlenebilir özelliğine sahiptir. İşlenebilen bu verilerden de anlamlı sonuçlar çıkarmak mümkün olmaktadır.

3-) Çeşitlilik (variety): Veriler sayısal, metinsel, görüntü, ses, video v.b farklı kaynaklardan ve farklı formatlardan elde edilerek üretilmektedir. Dolayısıyla veriler yapısal veya yapısal olmayan biçimde üretilmektedir (Erbay & Kör, 2016). Üretilen veriler, çoğunlukla yapısal olmayan ve farklı kaynaklardan farklı formatlarda elde edilen verilerdir. Büyük ölçekli ve farklı formatlardaki bu tür veriyi işlemek için ilişkisel veri tabanları yetersiz kalmaktadır. Dolayısıyla veri madenciliği, metin madenciliği, doğal dil işleme v.b sistemler kullanılarak yapısal olmayan veriler dönüştürülmekte, anlaşılabilir ve işlenebilir hale getirilmektedir. Ayrıca, yapısal olmayan verinin depolanabilmesini destekleyen ve dağıtık paralel işlem kabiliyetine sahip sistemlerin kullanımına da ihtiyaç duyulmaktadır (Zafar, vd. 2016, s.120).

4-) Doğrulama (verification): Veri elde ediniminde verinin doğru ve güvenilir kaynaktan elde edilmesi, güvenlik seviyesinin izlenmesi (Göksu, 2014), veriyi işleyen kişilerin güvenlik kriterine dikkat etmesi ve verinin belirli koşullar altında gizliliğinin sağlanması gerekmektedir. İşte karar mekanizmalarında karar alıcılara verinin güvenilebileceği ve kullanılabilceği ölçüde güvenilir olması gerekmektedir. Büyük verinin yüksek çeşitlilikte olması, analiz edilen verinin kalitesini ve güvenilirliğini sağlama sürecini zorlaştırmaktadır (Chandra, Ray, Goswami, 2017,

s. 49). Büyük verinin çok önemli bir boyutu olan gerçeklik veri kalitesi ile değerlendirilmektedir. Güvenilir bilgiler, yalnızca yüksek kaliteli veri ile üretilebilmektedir. Doğru veya geçerli olmayan büyük miktardaki veri, analizler için hem temel teşkil etmeyecek, hem de yanlış yorumlamalara yol açabilecektir (Gahi, Guennoun, Mouftah, 2016, s. 953).

5-) Değer (value): Büyük veri kurumların ve toplumun hedef ve beklentileri doğrultusunda organizasyonlar ve toplum için ekonomik değer oluşturmaktadır (Mauro, Greco ve Grimaldi, 2016, s. 131). Elde edilen verinin kullanılarak değere dönüştürülmesi büyük verinin 5. bileşenini oluşturmaktadır. Bundan birkaç yıl önceye kadar büyük hacimli verilerin hem üretim imkânının sınırlı olması hem de bunu işleyecek teknolojilerin olmaması nedeniyle derlenen verilerden istifade edilememekteydi. Büyük veri ile teknolojinin gelişmesi ve büyük verinin kullanılması ile ham veri, bilgiye dönüşerek yeni değer oluşturma fırsatı ortaya çıkmıştır. Büyük hacimli olan bu veriler bilgiye dönüştükten sonra işletmelerin iş süreçlerine ve karlılıklarına da katkı sunmaktadır.

Büyük veri geleneksel yöntem veya araçlarla işlenerek kullanılabilir hale getirilemeyen geleneksel analiz yöntemleri ile bilgisayarların işleyemeyeceği kadar büyük veriler anlamına gelmektedir. Büyük veri kavramına genel itibarıyla bakıldığında bir veya birden fazla kaynaktan farklı ortamlarda ve farklı formatlarda elde edilebilen ve sonuçlarından anlamlı yargılara ulaşılabilen büyük hacimli veri olarak ifade edilmektedir. Büyük verinin kullanımının temelinde yapılandırılmış veri üretmekten ziyade büyük hacimlerdeki yapılandırılmamış farklı ortamlardaki verilerden işlenerek anlamlı sonuçlar elde edilmesi yatmaktadır. İstatistikî yöntemler hem düzenli veriler üzerinde hem de düzenli olmayan büyük veri kapsamında toplanan düzensiz veriler üzerinde anlamlı sonuçlar elde etme imkânı sağlamaktadır.

Günümüzde büyük verinin sunduğu fırsatlar ile birlikte işletme yöneticileri ve sahiplerinin sezgiye ve tecrübeye dayalı karar verme sürecinden veriye dayalı karar verme süreçlerine doğru evrildiği bir dönem yaşanmaktadır. Dolayısıyla zaman içerisinde işletmelerde de veri odaklı yönetim kültürüne yönelim oluşacaktır. (Altunışık, 2015, s. 58). Büyük veri, verilerin dijitalleşmesi ve farklı boyutlarda alınan verilerin toplanıp, düzenlenerek insan davranışlarını anlama, tahminler yapma, öngörülerde bulunma (Tang & Karim, 2017, s. 35) işletmenin sahip olduğu verilerden yola çıkarak rasyonel ve akılcı yönetim imkânları sağlamaktadır. Dolayısıyla büyük veri kullanımı mevcutta kullanılan yazılım ve analiz araçları ile makul bir zaman diliminde analiz edilmesi mümkün olmayan büyüklükteki veri setlerini çözümlmeyi ve anlamlandırmayı ifade etmektedir. Büyük veri vasıtasıyla işletme yöneticileri ve sahipleri, hem operasyonel hem de taktik seviyede hızlı karar alabilmektedirler. Büyük veri ile aynı zamanda karşılıklarına çıkan problemlere cevap verme sürelerini hızlandırabilmektedirler (Altunışık, 2015, s. 58).

Elde edilen bütün veri kaynakları karşılıklı yarara dayalı ilişkilerin geliştirilmesi, insan davranış ve duygularının anlaşılması veya öngörülmesi ve kişiler arasında gerçekleşen ilişkileri ortaya koymak için kullanılabilir (Demirtaş & Argan, 2015, s. 5). Büyük veriler, veri setine uygun programlar aracılığıyla analiz edildiği takdirde insanların sergiledikleri davranışlar ve davranışların değişimlerinin daha yakından ve gerçek zamanlı olarak yönetilmesi

kolaylaşacaktır. Dolayısıyla büyük veriden sosyal bilgiler alanında da faydalanılması mümkün görülmektedir. Bununla birlikte büyük veri ile çalışmak; bilgisayar bilimleri, istatistik, sosyoloji, psikoloji, pazarlama v.b ilgili alandan oluşan birkaç farklı disiplinin bir arada çalışmasını gerektiren çok boyutlu bir yapıyı gerektirmektedir.

Veri Elde Ediniminde Geleneksel Yöntemler İle Büyük Verinin Karşılaştırılması

Veri toplama kaynaklarının birincil tür olan anket, deney ve gözlemden, ikincil tür olan hazır veriden elde edilen veri üretimine doğru evrilmesi sosyal bilimlerde de büyük verinin kullanımına yönelik ilgiyi artırmaktadır. Araştırmalarda büyük verinin kullanımı; geleneksel yöntemlere göre veri toplayan araştırmacılara kıyasla birçok avantaj sunmaktadır. Veri elde ediniminde geleneksel yöntemler ile büyük verinin karşılaştırılması Tablo1’de detaylı olarak sunulmaktadır;

Tablo 1. Veri Elde Ediniminde Geleneksel Yöntemler İle Büyük Verinin Karşılaştırılması

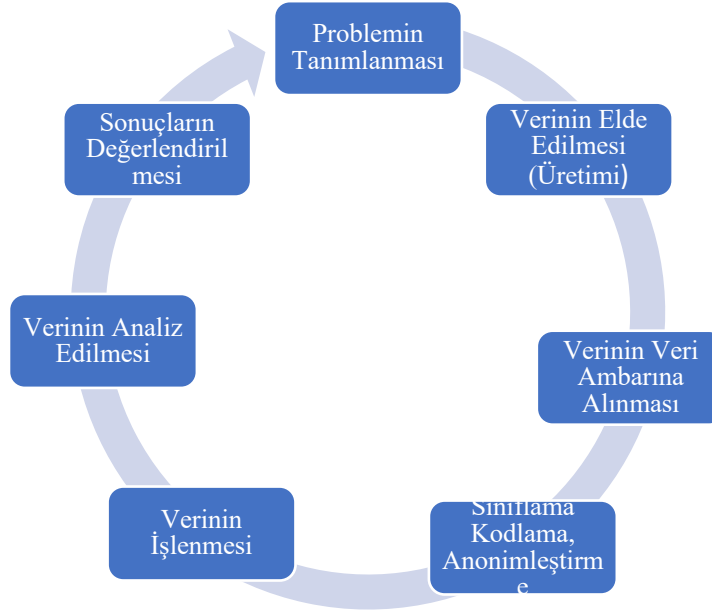
| Geleneksel Veri Toplama Yöntemleri | Büyük Veri |
|---|---|
| Veri toplama aşamasında elde edilmesi planlanan sonuca yönelik planlama yapılmaktadır. | Veri farklı platformlardan temin edilir, farklı amaçlar için toplanılan veri, mevcut projede ihtiyaca göre kullanılmaktadır. |
| Veri toplamada belirli kurallar ve varsayımlar bulunmaktadır. | Veri toplama süreci tam olarak yapılandırılmamıştır. |
| Soru kağıdı kullanımı ve çalışmaya özel soru tasarımları gerektirmektedir. | Elde edilen görüntü, ses, arama kayıtları, loglar, sensörler, v.b bütünleştirilmiş veri ile analizler yapılmaktadır. |
| Cevaplayıcılarla yüz yüze görüşmeler yapılmakta ve bu görüşmeler cevaplayıcılara yük oluşturmaktadır. | Cevaplayıcı ile muhatap olunmamakta buna karşın ilgili kişilerin bilgileri doğrudan kullanılmaktadır. |
| Verisi toplanan 3. kişilerden onay (olur) alınmaktadır. | Verisi toplanan 3. kişilerden onay (olur) varsayımsal olarak alınmaktadır. |
| Saha araştırması ve bileşenleri (anketör, ulaşım, v.b.) nedenlerle maliyetler yüksektir. | Bilgi teknolojilerine ilk yatırım maliyeti yüksek olmakla birlikte sonraki maliyetler göreceli olarak daha düşüktür. |
| Tam sayımın gücü nedeniyle örnekleme kullanımını gerektirmektedir. | Örnekleme yöntemi kullanılmamakla birlikte bütün veri üzerinden analizler yapılmaktadır. |
| İstatistik analizlerinde uzmanlaşmış personellere ihtiyaç bulunmaktadır. | Birkaç bilgisayar programını beraber kullanabilen, veri bilimcisine (data scientist) ihtiyaç bulunmaktadır. |
| SPSS, SAS gibi paket programlar aracılığıyla elde edilen veriler analiz edilmektedir. | Hadoop bileşenleri gibi yoğun bilgi işlem teknolojilerinin kullanımını gerektirmektedir. |
| Geçmiş zamana dönük elde edilen veriler analiz edilmektedir. | Çalışmanın türüne göre sensörlerin de yardımıyla geçmişe dönük ve anlık, gerçek zamanlı (real time) olay kayıtları (loglar) üretilebilmektedir. |
| Veri gizliliği riski kısmen düşüktür. | Veri gizliliği riski oldukça yüksektir. |
| Veri miktarına odaklanılmaktadır. | Veri akışına odaklanılmaktadır. |
| Yapısal, yarı yapısal ve çoğunlukla sayısal veriler kullanılmaktadır. | Çoğunlukla sayısal veriler yapılandırılmış ve yapılandırılmamış her tür veri (resim, tekst, sayılar, indeksler) kullanılmaktadır. |

Kaynak: Abreu ve Acker (2013). Context and collection: A Research Agenda for Small Data, Conference 2013 Preceedings, February 12-15, Fort Worth, TX, USA’dan uyarlanarak genişletilmiştir.

Tablo 1’de görüldüğü üzere geleneksel veri toplama yönteminden büyük veriye geçiş sürecinde veri toplama süreci, örnekleme yöntemi, analiz yöntemi, veriyi analiz eden personelin niteliği, bilgi işlem süreci, veri toplanan grubun nitelikleri farklılıklar göstermektedir.

Araştırmalarda Büyük Veri Kullanım Döngüsü Nasıl Gerçekleşir?

Büyük veri çalışma alanı yeni ve hızlı biçimde gelişmektedir. Genellikle yapısal olmayan veriler üzerinden çalışmalar gerçekleştirildiği için bu çalışmalarda büyük veri kullanım döngüsüne ilişkin genel bir model oluşturmak gerçekten zordur. Bununla birlikte Essex Üniversitesi verinin yaşam döngüsü (Corti, Eynden, Bishop, & Woollard, 2014) ile EMC firması tarafından geliştirilen büyük veri yaşam döngüsü (Schmarzo, 2012) modellerinden uyarlanarak hazırlanan büyük veri kullanım döngüsü Şekil.1’de özetlenmektedir.



Şekil 1. Büyük Veri Kullanım Döngüsü

Kaynak: Şeker, Sadi Evren, Büyük Veri ve Büyük Veri Yaşam Döngüleri YBS Ansiklopedi www.YBSAnsiklopedi.com, Cilt 2, Sayı 3, Eylül 2015, Essex Üniversitesi Verinin Yaşam Döngüsü ile EMC Firması Büyük Veri Yaşam Döngüsü Modellerinden uyarlanmıştır.

Büyük veri kullanım döngüsü Şekil 1’de görüldüğü üzere; problemin tanımlanması, veri elde edilmesi, verinin veri ambarına alınması, sınıflama, kodlama, anonimleştirme, verinin işlenmesi, verinin analiz edilmesi, sonuçların değerlendirilmesi olmak üzere 7 aşamadan oluşmaktadır. Aşağıda bu aşamalarda gerçekleştirilen işlemler detaylı olarak sunulmaktadır.

Problemin Tanımlanması

Her araştırmada olduğu gibi ilk önce araştırma yapılacak olan konunun belirlenmesi ve problemin tanımlanması gerekmektedir. Bu bölümde özellikle üzerinde durulması gereken konu; gerçekleştirilecek olan araştırma, büyük veri kullanımını gerektirmekte midir? Ayrıca ilgili konuda veri sağlama ve temin yolları açık mıdır? Problemin tanımlanması aşamasında büyük veriden elde edilecek sonuçlar çözüm için yardımcı olup olmayacağı önceden tahmin yoluyla kestirilmelidir.

Verinin Elde Edilmesi (Üretimi)

Verinin elde edilmesi sürecinde ilk olarak araştırmancının amacına bağlı olarak eldeki mevcut kaynaklardan elde edilen verinin bir araya getirilmesi gerekmektedir. Sonrasında alternatif veri kaynaklarının neler olduğunun araştırılması; kamuda idari kayıtlar ile özel sektörde araştırma yapılacak konu hakkında veri olup olmadığı, belirli protokoller çerçevesinde verinin kullanılıp kullanılmayacağı araştırılmalıdır. Başka bir ifadeyle belirlenen konuyla ilgili resim, video, idari kayıt, güncel raporlar, makaleler, sayısal veriler v.b tüm yapısal ve yapısal olmayan veri kaynakları bir araya getirilmelidir. Veri elde edinimine yönelik karşılıklı mutabakata varılmasından sonra gizlilik sözleşmeleri imzalanmalıdır. Burada dikkat edilmesi gereken konu verinin olabildiğince amaca hizmet edecek şekilde güncel ve güvenilir olması, kapsamının geniş olması ve veride eksiklik olmamasıdır.

Verinin Veri Ambarına Alınması

Elde edilen verilerin üretim amacı, niteliği, büyüklüğü, fiziki ortamlarda bulunma yapısı ve çeşitliliği farklılık gösterdiği için bu bölümde temel olarak verinin alınıp diğer verilerle entegre edilmesi söz konusudur.

Sınıflama, Kodlama, Anonimleştirme

Verinin sınıflanması aşamasında farklı amaçlarla derlenen verilerin belirli bir amaca hizmet edebilmesi için ortak bir dil olarak sınıflandırılması gerekmektedir. Verileri işlemeyi kolaylaştırmak adına, verilerin amaca uygun şekilde daha yalın sembollerle ifade edilmesi kodlama olarak ifade edilmektedir. Bu aşamada elde edilen verinin birbiriyle konuşabilmesi için kodlanması gerekmektedir. Verinin anonim hale getirilmesi veri kümelerinde, dolaylı tanımlayıcıların doğru kombinasyonlarla bir araya gelmesi halinde kayıtlardaki kişilerin kimliklerinin belirlenememesi veya belirli bir kişiye dair bilgilerin rahatlıkla tahmin edilemeyecek duruma getirilmesidir (KVKK, 2017, s. 28). Dolayısıyla araştırmada verisi kullanılan 3. kişilerin detaylı şahsi bilgileri tanımlamaya yol açmayacak şekilde anonimleştirilmesi gerekmektedir.

Verinin İşlenmesi

Sınıflama, kodlama ve anonimleştirme işlemlerinin ardından veri işleme aşamasına geçilmektedir. **Veri işleme**, tamamen ya da kısmen otomatik yollarla veya herhangi bir veri kayıt sisteminin parçası olmak kaydıyla otomatik olmayan yollarla verilerin ilk defa elde edilmesinden başlayarak veriler üzerinde gerçekleştirilen işlem türlerini ifade etmektedir. Bu işlemler toplama, kaydetme, paylaşma, depolama, okuma, değiştirme, açıklama, imha etme gibi birçok alandan oluşmaktadır (KVKK). Bu aşamada elde edilen veri üzerinde veri madenciliği, metin madenciliği, doğal dil işleme v.b sistemler kullanılarak yapısal olmayan veriler dönüştürülmekte ve böylece anlaşılabilir ve işlenebilir hale getirilmektedir. Verinin işleme aşamasında verinin akışkan olup olmadığı, sürekliliğinin sağlanması, düzensiz verilerin ayıklanıp düzenli hale getirilmesi için hangi yapısal temellerin oluşturulması gerektiği tespit edilmelidir.

Verinin Analiz Edilmesi

Veri analiz sürecinde en önemli konu, elde edilen verilere hangi yöntem ve tekniklerin uygulanacağıdır. Bu aşamada veri bilimciler ekonometrik modeller kurmaya ve kurulan modellerin doğrulamasına çalışarak belirlenen probleme ilişkin çözüm yolları aramaktadırlar. Standardize olmayan verilerden belirli tekrarlar, belirli şablonlar yakalanmaya çalışılmaktadır. Bu aşamada, görüntü ve uzaysal veriler, günlük dosyalar, arama kayıt verileri, makine ve akıllı okuyucu verileri gibi yapısal olmayan veri kaynakları ilişkisel olmayan veri platformuna Hadoop ekosistemine aktarılmaktadır. HDFS, Hive, Hbase gibi Hadoop bileşenleri, veriyi düşük maliyetle saklamak ve yönetmek, MapReduce ve Pig anlamlandırmak, YARN kaynakları yönetmek ve diğer Hadoop araçları da bütün veri analizinde eksik kalan parçaları tamamlamak için kullanılmaktadır (Özdoğan, 2016, s. 35).

Sonuçların Değerlendirilmesi

Elde edilen sonuçlar, iş zekâsı ileri analitik teknikleri, görüntüleme ve araştırma sonuçları (Aktan, 2018, s. 11) ile görsel sunumlarla ve raporlamalar ile farklı formatlarda kullanıcılara sunulabilmektedir. Aynı zamanda, büyük veri kullanılarak elde edilen sonuçlar literatürdeki örneklem yoluyla gerçekleştirilen diğer araştırmalarla elde edilen sonuçlar ile mutlaka karşılaştırılmalıdır.

Büyük veri kullanım döngüsü incelendiğine verinin temin edilmesinden sonuçların değerlendirilmesine kadar yoğun bir süreç olduğu görülmektedir. Matematik, İstatistik, Bilgisayar bilimlerinin bir arada kullanabilen veri analistlerine büyük veri kullanım döngüsü sürecinde ihtiyaç duyulmaktadır. Büyük veri kullanım döngüsü içerisinde kullanılan, üzerinde düzenleme yapılan ve çıktı olarak elde edilen veriler hiç şüphesiz başka araştırmalarda da girdi olarak kullanılabilir.

Büyük Verinin Sosyal Bilimlerde Kullanım Alanları

Büyük verinin kullanım alanlarını açıklayabilmek için öncelikle nesnelerin interneti kavramı üzerinde durulması gerekmektedir. Nesnelerin interneti; nesnelerin çeşitli bağlantılarla internete erişip diğer nesnelerle iletişim halinde olmasıdır. İnternete bağlı olan tüm nesneler ağ geçitleri aracılığıyla bir ağa bağlanmakta, sonrasında bu nesnelerin durumuna göre tasarlanan ısı, ses, koku, ağırlık gibi fiziksel özellikleri algılayan sensörler aracılığıyla toplanan veriler; ağ sayesinde bulut (online depolama hizmeti) platformuna ulaşıp bu verinin aktarımıyla ve verinin işlenmesiyle bir değer oluşturulmaktadır. Oluşan bu değer algoritmalarla şifreledikten sonra kullanıcının sistemle entegre olan cep telefonu, bilgisayar gibi cihazlarına programlar ve uygulamalar aracılığıyla ulaştırılmaktadır (Öztopal, 2018). Akıllı saatler, akıllı telefonlar, akıllı ev aletleri, yeni nesil otobüs durakları ve daha birçok konu nesnelerin internetine örnek gösterilebilmektedir. Nesnelerin internetinin sunduğu veri üretme fırsatı nedeniyle günümüzde neredeyse her alet internete bağlı olacak şekilde tasarlanmaktadır. Dolayısıyla tasarlanan nesnelerin hareketleri internet ortamında kaydedilmekte ve karar vericiler için bu kayıtlar bir değer üretmektedir.

Büyük verinin kullanım alanı araştırmacıların biraz da hayal gücünün genişliği ile de ilişkilidir. Büyük veri, sosyal bilimlerde de birçok alanda kullanılabilir. Büyük veri özellikle

işletmelerde insan davranışlarını önceden tahmin etme olanağı sağlayarak karar alıcı olan yöneticilere fırsatlar sunabilmektedir. Örneğin; belirli zaman diliminde, imalat işletmelerinde çıktı oluşturabilecek üretim miktarları ile çalışanların işe giriş çıkış saatleri, ilgili işletmelerde çalışanların sosyal medya hesaplarındaki paylaşımları arasındaki ilişkiler dikkate alınarak örgütsel davranışın çalışma alanlarına giren tükenmişlik, sinizm, işe yabancılaşma, duygusal zekâ, örgütsel bağlılık, iş tatmini v.b konular ile üretim miktarı arasında ilişki olup olmadığına bakılabilir.

Hastanelerin sağlık kayıtları sektör bazında (madencilik, inşaat, gıda, finans, tekstil, ticaret, turizm v.b) işyerleri ile birleştirilip ülke, bölge ve il düzeyinde meslek hastalıklarına ilişkin büyük veri sonuçları kullanılabilir ve karşılaştırılabilir. Burada T.C kimlik numarası işyerleri ve sağlık kayıtları arasında anahtar kimlik olarak eşleştirmelerde bağlantı rolünü gerçekleştirebilecektir.

Bir başka araştırmada; çalışanların iş ortamında internette kalma süreleri, araştırma yaptığı siteler ile iş tatmini, işten ayrılma niyeti v.b konular arasında ilişki olup olmadığına bakılabilir.

Büyük verinin kullanılacağı her ortamda kişinin dolaylı olarak tanımlanabileceği mikro veri detayına inmek hukuksal sorunlara ve problemlere neden olacaktır. Bu sebeple yukarıda sosyal bilimler alanında örnekleri verilen yapılabilecek çalışmaların toplu veri üzerinden kişiyi tanımlamayacak düzeyde ve makro bazda analizler yapılarak gerçekleştirilmesi gerekmektedir.

Büyük Veri Kullanımının Dezavantajlar Ve Dikkat Edilmesi Gereken Hususlar

Büyük veri ile ilgili teknolojinin yeni olması hazır teknolojilerin bu yapıyla uyumsuzluğu, yatırım maliyetinin yüksek olması, yabancı teknoloji şirketlerine kaynak bağımlılığının yüksek olması (Altunışık, 2015, s. 47) bu yeni teknolojinin yararlanıcıları için büyük avantajlar yanında dezavantajlar oluşturmaktadır. Aşağıda büyük veri kullanımının dezavantajlarından bazıları sıralanmıştır;

Büyük veri çalışmaları genellikle açık veri platformlarında gerçekleştirildiği için analize dahil edilen verilerin programın üreticileri tarafından elde edilmesi veri gizliliği ve güvenliliği açısından büyük risk taşımaktadır. Bu sebeple güvenli bir ortam sağlamak için büyük veri kullanılabilir yerli yazılımların geliştirilmesi gerekmektedir.

- Büyük veri çalışma ortamı açık kaynak kodlu yazılım kullanılması ve bu ortama 3. kişilerin saldırısı nedeniyle ortamdaki bilgilerin çalınması söz konusu olabilmektedir.
- Vatandaşların günlük yaşantılarına getirdiği potansiyel mahremiyet ve güvenlik sorunlarına ilişkin henüz etkin çözümler geliştirilememiştir. Çözüm önerisi olarak verisi kullanılan kişilerin veriyle bağlantısını tanımlamayacak şekilde kukla değişkenler kullanılması gerekmektedir.
- Nesnelerin interneti kapsamında veri üreten nesne ve cihazların birçoğu her zaman internete bağlı haldedir ve dolayısıyla her zaman saldırıya açıktır.

- Büyük veriye ilk yatırım maliyetleri oldukça yüksektir. Bu sorunu bulut tabanlı teknolojilerin gelişmesi ile birlikte saatlik veya günlük maliyetleri karşılayabilme olanağı bireysel veya kurumsal kullanıcılara sunulmaktadır. Böylece daha az maliyetle büyük miktarda veriler işlenebilmektedir.

Sonuç ve Tartışma

Sosyal bilimlerde araştırmacıların araştırma yaptıkları konularda zorlandıkları süreçlerden birisi veri toplama süreçleridir. Veri toplama süreçlerinin birincil kaynaklardan ikincil kaynaklara yöneldiği günümüz dünyasında büyük veriyi anlamak büyük veriden anlamlı veriler üretmek Sosyal Bilimlerde araştırmacıların ilgisini çekmektedir. Büyük verinin kullanımının avantajlarına bakıldığında;

- Örneklem yöntemine göre daha geniş kapsamlı büyük veriden anlamlı sonuçlar elde etmek mümkündür,
- Veri toplama sürecinin zaman ve maliyet yönüyle zahmetli olmasına karşın büyük veriyi elde etmek kısmen daha kolaydır,
- Gerçek zamanlı (real time) veri üretimi mümkündür, başka bir ifade ile verinin elde edilmesiyle analiz edilip yayınlanması arasında geçen süre oldukça kısadır,
- Büyük veri ile daha güncel ve gerçek zamanlı istatistiklerin kullanımı mümkün olmaktadır.
- Büyük veri kullanımı ve elde edilmesindeki zorluklar ve dezavantajlarına bakıldığında;
- İlk yatırımdaki bilgi işlem maliyetleri oldukça yüksektir,
- Büyük veri ile elde edilen verilerde analiz ve işlemlerin önem kazanmasıyla birlikte elde edilen veriyi analiz edebilecek nitelikli personel ihtiyacı artmıştır,
- Veri üretiminde süreklilik ve düzenliliğin, akışkanlığın zaman serilerinin sağlanması gerekmektedir,
- Veri gizliliği ve mahremiyetini sağlamak oldukça zordur.

Sürekli değişen ve gelişen koşullarda sosyal bilimlerde içinde sürdürülebilir rekabet üstünlüğü sağlamanın kaynaklarından birisi farklılaştırma stratejisinin uygulanmasıdır. Bu noktada büyük verinin kullanımı araştırmacılara farklı ve yeni yol, yöntem ve süreçlerin ortaya çıkarılmasında yardımcı olabilecek çalışmalara ivme kazandırabilecektir.

Büyük veri konusunda ilerleme sağlayabilmek için veri analizi becerisine sahip veri bilimcilerin yetiştirilmesi, kamu-üniversite-özel sektör işbirlikleri ulusal ve uluslararası düzeyde mutlaka yapılmalı ve deney laboratuvarlarının, araştırma geliştirme merkezlerinin kurulması için yatırımlar yapılması gerekmektedir. Ayrıca büyük veri hakkında hukuki altyapının da zaman içerisinde hazırlanarak yürürlüğe konması gerekmektedir.

Kaynakça

- Abreu, A., & Acker, A. (2013). *Context and collection: A research agenda for small data*. iConference 2013 Proceedings, 549-554, doi:10.9776/13275.
- Aktan, E. (2018). Büyük veri: uygulama alanları, analitiği ve güvenlik boyutu. *Bilgi Yönetimi Dergisi*, 1(1).
- Altunışık, R. (2015). Büyük veri: fırsatlar kaynağı mı yoksa yeni sorunlar yumağı mı? *Yıldız Social Science Review*, 45-76.
- Chandra, S., Ray, S. ve Goswami, R. T. (2017). *Big data security: survey on frameworks and algorithms*. 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, 48-54.
- Corti, L., Eynden, V. V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data*. A Guide to Good Practice. Essex: SAGE.
- Demirtaş, B. & Argan M. (2015). Büyük veri ve pazarlamadaki dönüşüm: kuramsal bir yaklaşım. *Pazarlama ve Pazarlama Araştırmaları Dergisi*, 15(Ocak), 1-21.
- Diebold, W. Francis (2000). *Big data dynamic factor models for macroeconomic measurement and forecasting, advances in economics and econometrics*, Eighth World Congress of the Econometric Society. Cambridge: Cambridge University Press, 115-122.
- Doğan, K. & Aslantekin, S. (2016). Büyük veri: önemi, yapısı ve günümüzdeki durum. *DTCF Dergisi*, 56(1), 15-36.
- Erbay, H. & Kör, H. (2016). *Büyük veri ve büyük verinin analizi*. Uluslararası Bilim ve Teknoloji Konferansı 3-6 Ekim Ankara.
- Gahi, Y., Guennoun, M., Mouftah, H. T. (2016). *Big data analytics: security and privacy challenges*. 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, Italy, 952-957.
- Göksu, C. (2014). Datawarehouse Türkiye. (2019, Ekim, 12) Erişim adresi: <http://datawarehouse.gen.tr/big-datanedir-geleneksel-veri-yonetimine-etkisi-ne-olur/>.
- Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J., Ramakrishnan, J.M.R. ve Shahabi, C. (2014). Big data and its technical challenges, *Communications of the ACM*, 57 (7), 86-94.
- KVKK, (2017). Kişisel verilerin silinmesi, yok edilmesi veya anonim hale getirilmesi rehberi. Kasım, 1-57.
- KVKK, (2019). Veri İşleme nedir? (2019, 15 Ağustos) Erişim adresi: <https://www.kisiselverikanunu.com/veri-isleme-nedir/>.
- Mathai, Paul (2019). Big Data: Catalysing performance in manufacturing. Applied Research Lead, Manufacturing & Hi-Tech, Wipro, (2019, 7 Haziran) Erişim adresi: <https://www.wipro.com/content/dam/nexus/en/industries/process-and-industrial-manufacturing/latest-thinking/2606-Big%20Data%20-%20Copy.pdf>.
- Mauro, A. D., Greco, M., Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, 65(3), 22-135.
- Özdoğan, O. (2016). *Büyük veri denizi*. Ankara: Elma.

- Öztopal, B., Dünyamız İçin Yeni Bir Adım: Nesnelerin İnterneti. (2019, 10 Temmuz) Erişim adresi: <https://medium.com/d%C3%B6n%C3%BCm-noktas%C4%B1/d%C3%BCnya-i%CC%87%C3%A7in-yeni-bir-seviye-nesnelerin-i%CC%87nterneti-ac726fb95b71>.
- Rubistein, I.S. (2013). Big data: the end of privacy or a new beginning? *International Data Privacy*, 3(2), 4- 86.
- Sağiroğlu, Ş. (2017). *Açık veriler: büyük veri büyük etki*. Açık Veri Konferansı, 23 Mayıs 2017, Gazi Üniversitesi, Ankara.
- Sağiroğlu, Ş. & Sinanç, D. (2013). *Big data: a review*. International Conference on Collaboration Technologies and Systems (CTS), (2013, 20-24 May), San Diego, CA, U.S.A., 42-47.
- Schmarzo, B. (2012). New Roles In The Big Data World. *EMC, Infocus*.
- Şeker, S. E., (2015). *Büyük veri ve büyük veri yaşam döngüleri YBS Ansiklopedi*, 2(3).
- Tang, J. J. & Karim, K. E. 2017. Big data in business analytics: implications for the audit profession. *CPA Journal*, 87(6), 34-39.
- Turan, M., (2016). *Dijital çağda büyük veri*. Türkiye Kalkınma Bankası Yayını, Ocak - Mart 2016: 79.
- Vinod, B. (2013). Leveraging big data for competitive advantage in travel. *Journal of Revenue and Pricing Management*, 12(1), 96-100.
- Zafar, R., Yafi, E., Zuhairi, M. F. ve Dao, H. (2016). *Big data: the nosql and RDBMS review*. International Conference on Information and Communication Technology (ICICTM), (2016, 16-17 May) Kuala Lumpur, Malaysia, 120-126.