

**İŞGÜCÜ PİYASALARININ İZLENMESİNDE BÜYÜK VERİ GOOGLE TRENDS  
VERİLERİ ÜZERİNDEN BİR ARAŞTIRMA****BIG DATA MONITORING IN LABOR MARKETS A RESEARCH ON GOOGLE TRENDS  
DATA****Doç. Dr. Barış SEÇER**Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Çalışma Ekonomisi ve Endüstri  
İlişkileri Bölümü Öğretim Üyesi, İzmir/Türkiye**Doç. Dr. H. Şebnem SEÇER**Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Çalışma Ekonomisi ve Endüstri  
İlişkileri Bölümü, Öğretim Üyesi**ÖZ**

Büyük veri uygulamalarının işgücü piyasalarının izlenmesinde kullanılmasının önemi yazında vurgulanmaktadır. Büyük veri, işletmelerin işgücü talebi hakkında farklı bilgiler edinilmesini sağlayabilmektedir. İşletmelerin internet ortamındaki iş ilanlarından nitelik talepleri, cinsiyet ayrımcılığı gibi konular araştırılabilmektedir. Ayrıca işsizliğin izlenmesinde geleneksel veri kaynaklarına ek kaynakların ortaya çıktığı görülmektedir. Özellikle son dönemde Google arama verilerine dayalı olarak işsizliği inceleyen çok sayıda araştırma göze çarpmaktadır.

Google arama verilerinin ülkemizde işsizliğin tahmininde etkisini belirlemek üzere bir araştırma yapılmıştır. Çalışmada 15 ve üstü yaş ile 15-24 yaş grubunun işsizlik oranları ve 4 Google arama terimi incelenmiştir. İşsizlik oranını tahmin etmek için ARIMA modellerinden yararlanılmıştır. Bulgulara göre, gerçekleşen ve tahmin edilen işsizlik değerlerinin birbirine yakın olduğu görülmektedir. Kısaca Google Trends verilerinin işsizliğin tahminlenmesinde katkıda bulunabileceği anlaşılmaktadır.

**Anahtar Kelimeler:** Büyük veri, işsizlik, işgücü piyasaları, işsizlik tahmini

**ABSTRACT**

Emphasis is placed on the importance of using big data applications in monitoring labor markets. Big data can make it possible for enterprises to obtain different information about labor demand. Job advertisements on the internet can be used to investigate issues such as quality requirements and gender discrimination. In addition, additional resources are emerging for traditional data sources when unemployment is monitored. Especially in the recent period, based on Google search data, there are a lot of researches that investigate unemployment.

A search has been made to determine the impact of Google search data in estimating unemployment in our country. Unemployment rates for 15 and over age group and 15 to 24 age group and 4 Google search terms were examined in the study. ARIMA models were used to estimate the unemployment rate. According to findings, it is seen that actual and estimated unemployment values are close to each other. In short, it appears that Google Trends data may contribute to the prediction of unemployment.

**Keywords:** Big Data, Unemployment, Labor Market, Unemployment forecast

**1. GİRİŞ**

İnsanlığın yüzyıllardan beri ürettiği veri miktarından daha fazlasının gelişen teknoloji sayesinde çok daha kısa bir sürede üretildiği bir dönemde olduğumuz belirtilmektedir. Bu dönemde cep telefonları, internet, sosyal medya gibi araçlar çok miktarda verinin toplanmasına yardımcı olmaktadır. Çok miktarda veri aynı zamanda depolama, işleme ve analiz açısından da yeni sorunlar ve fırsatlar sunmaktadır. Bu bağlamda, büyük verinin karar alma sürecine önemli etkisi olacağı, kararların daha fazla veriye dayalı olarak alınmasının mümkün olacağı savunulmaktadır. Benzer şekilde işgücü piyasasının izlenmesi açısından da etkileri olacağı yazında açıklanmaktadır.

Çalışmada öncelikle büyük veri kavramı, özellikleri ve etkileri açıklanacaktır. Ardından işgücü piyasasında büyük verinin etkilerine değinilerek, önce işgücü talebinin ölçüsülmesine sağladığı katkılar, sonrasında ise

işsizliğin öngörülenmesine getirdiği yenilikler açıklanacaktır. Son olarak Google arama verilerinin ülkemizde işsizliğin tahmininde kullanılıp kullanılmayacağını belirlemek üzere yapılmış araştırma sonuçlarına yer verilecektir. Araştırmada 15 ve üstü yaş ile 15-24 yaş grubunun işsizlik oranları ve 4 Google arama terimi ekonometrik modellerle incelenmiştir. İşsizlik oranını tahmin etmek için ARIMA modellerinden yararlanılmıştır.

## 2. BÜYÜK VERİ

### 2.1. Tanım

Büyük veri akıllı telefonlar, web siteleri, uygulamalar, algılayıcılar, barkodlar ve sosyal medya mesajları gibi farklı sayısal araçlar tarafından üretilen veriyi ve bu verinin analiz edilmesini içermektedir. Büyük veri ticaret, hukuk ve bilişim sistemleri gibi farklı alanlarda farklı yönleriyle tanımlanmaktadır. Bu anlamda büyük verinin tek boyutlu bir kavram olmadığı ileri sürülebilir. Gerçekten büyük veri kavramı teknoloji, analitik yöntemler, iktisat, ticaret ve hukuk açısından tanımlanabilmektedir.

Bazı yaklaşımlar veri depolama ve analiz yöntemlerinin günümüz kültürü ve zamanından ileri olmasından dolayı büyük kavramının kullanıldığını belirtmekte ve bu yönde tanımlar yapmaktadır. McKinsey&Company (2011) danışmanlık şirketi büyük veriyi, klasik veritabanı yazılımları aracılığıyla edinilemeyen, depolanamayan ve yönetilemeyen veri kümesi olarak açıklamaktadır. Ayrıca Page ve arkadaşları (2012), büyük veri kavramını enformasyonu analiz etme, yönetme, dağıtma, saklama ve elde etme olanağı veren gelişmiş teknikler ve teknolojiler gerektiren büyük miktarda hızlı, karmaşık ve değişken veri olarak değerlendirmektedir.

Öncelikle büyük veri standardına uyan veri miktarının teknolojik gelişimle değişebildiği ve artabildiği anlaşılmaktadır. Ayrıca farklı uygulamalarda büyük veri standardını karşılayan veri kümelerinin birbirinden farklı olduğu anlaşılmaktadır. Bu nedenle büyük veri için tek ölçüt veri kümesinin büyüklüğü değildir (Chen vd.,2014:5). Ayrıca veri türlerinin çeşitliliği, hız ve veriden çıkarımda bulunma süreci de vurgulanmaktadır.

Örneğin DeRoos (2012) büyük veriyi, “muazzam hacim, çeşitlilik ve hızdaki veriden, daha önce mümkün olanın ötesinde bir içgörü çıkarma” olarak tanımlamaktadır. Gerçekten büyük veri algoritmaları, ticari değeri olabilecek örneğin hava ve tüketim veya sağlık ve kredi riski arasındaki ilişkilerin ortaya çıkarılmasında yardımcı olmaktadır. Hatta terörizmle ilgili suçların araştırılması süreci büyük miktarda davranışsal verinin analiz edilmesine dayalı hale gelmiştir (Helbing, 2015:74).

Büyük verinin özelliklerini vurgulayan tanımlar da yapılmaktadır. Buna göre, büyük veri mevcut veya geleneksel yöntemlerin ve sistemlerin kullanılması ile önceden çözülemeyen sorulara özgün yaklaşımları mümkün kılan büyük, hızlı ve çeşitli sayısal veriden bahsetmektedir (Lake ve Drake,2014:2)

### 2.2. Özellikleri

Yukarıdaki tanımlara göre, büyük verinin büyüklük, çeşitlilik, hız ve kalite (doğruluk) şeklinde dört özelliği olduğu anlaşılmaktadır. Büyüklük, büyük miktarda verinin kaydedilmesi, analiz edilmesi ve yönetilmesidir. Veri büyüklüğü, kaynakların sayısı ve veri derinliğiyle artmaktadır. (King, 2013:34). İşlem kapasitesinin katlanarak artması dünya çapında veri miktarının artması ile ilişkilidir. Araştırmalara göre, veri miktarı iki yılda bir katlanmaktadır. Bu yapı, büyük veri düşüncesinin temeli sayılmaktadır. Sürekli yeni yaşam alanlarına ilişkin açıklamalar ve daha fazla verinin analizde kullanılması temel düşüncedir (Dorshel ve Dorshel, 2015:7).

Veriler, yapıları önemli ölçüde farklılaşacak şekilde, hatta şimdiye kadar bilinmeyen biçimlerde örgüt içi ve dışından yeni kaynaklardan gelmektedir. Çeşitlilik, sosyal medya içeriklerinden video akışlarına kadar bir alanı kapsamaktadır. Büyük veri ile ilgili verilerin çoğunluğu daha önceden tanımlanmış bir veri modeline uymayan yapılandırılmamış biçimdedir. Anadilde metinler ve ses verileri örnek olarak verilebilir (Dorshel ve Dorshel, 2015:8).

Verilerin üretim hızı değişmektedir. Bu durum hızlı analiz yapma ve karar almayı gerekli kılmaktadır. Hız, kaynakların miktarı ile veri üreten araçların artan işlem kapasitelerinden etkilenmektedir (King, 2013:35).

Doğruluk, verilerin kalitesi ve kaynağı ile ilgilidir. Verilerin kalitesi tutarsızlık, eksiklik ve belirsizliğin etkisi altındadır. Verilere dayalı kararlar kapsayıcılık ve temellendirebilirlik gerektirmektedir (King, 2013:35). Bu özellik, veri içeriğinin doğruluğu, bütünlüğü ve güvenilirliği ile ilgilidir. Büyük veri kullanımının bazı biçimlerinin nesnel bilgi değerinin ölçümü güvenilir değildir. Örneğin sosyal ağlarda kullanıcı tarafından oluşturulan metinler öznel duygular ile farklı zamansal ve içerik bağlamlarında şekillenmektedir. Büyük veri

ile planlama, uygulama ve değerlendirme yaparken bu faktörler dikkate alınmalıdır (Dorshel ve Dorshel, 2015:8).

### 2.3. Etkileri

Büyük verinin kararların farklı şekilde alındığı ve farklı şekilde ölçüldüğü bir dönemi başlattığı ileri sürülmektedir. 17. yüzyılda mikroskopun keşfedilmesiyle bilim adamlarına yeni imkânların doğması gibi, büyük verinin de dünyayı gözlemekte yeni bir dönem açtığı belirtilmektedir. Bu nedenle hiçbir ülkenin büyük veri potansiyelini göz ardı etmesinin mümkün olmadığı ve büyük verinin toplumdaki tüm sektörleri etkileyeceği ileri sürülmektedir.

Örneğin akıllı ölçme sistemleri ile enerji üretimi ve tüketiminin daha iyi eşleşebileceği, enerji kısıntılarının en aza indirilebileceği belirtilmektedir. Dolayısıyla kaynaklar daha etkin yönetilebilecek ve çevre daha iyi korunabilecektir. Ayrıca riskler daha iyi anlaşılabilir ve kaçınılabilir hale gelecek, böylece alınan kararların istenmeyen sonuçları azaltılabilecektir. Tıp açısından ise, hastalıkların önlenmesinin tedavi edilmesinden daha önemli hale geleceği ileri sürülmektedir. Büyük verinin işletmelerin faaliyetlerini önemli şekilde etkileyeceği ve sundukları hizmetleri dönüştüreceği de ileri sürülmektedir.

Sigorta firması The Climate Corporation, çiftçilere algoritmalara ve büyük veri analizine dayanan yeni bir sigorta türü sunmaktadır. Firma hava koşullarını her bir dönüm için büyük miktarda veriye dayanarak analiz etmekte ve sonra bu verileri hasat ve toprak verileri ile birleştirmektedir. Firma sadece hava koşulları için bir sigorta sunmamakta, aynı zamanda çiftçilere verimliliklerini nasıl arttıracakları ve sigorta taleplerini nasıl asgariye indirecekleri konusunda tavsiyelerinde bulunarak sigortacılık rolünü de değiştirmektedir. (Klous ve Wieleard, 2016:19)

Büyük veri, yeni uygulamaların ortaya çıkmasının yanısıra insanların davranışlarının da değerlendirilmesini sağlayabilmektedir. İsveç Karolinska Enstitüsü araştırmacıları 200.000 kişinin hayatını kaybetmesine neden olan 2010 Haiti depremi öncesi ve sonrasında insanların hareketlilik biçimlerini incelemişlerdir. Haiti'nin en büyük mobil iletişim şirketinden elde edilen veriler, depremden önceki 42 gün ile 158 gün sonrası yapılan 2 milyon telefon görüşmesinden oluşmaktadır. Telefon görüşmeleri verilerinden deprem günü Port-au-Prince şehrinde bulunan 630.000 kişinin 3 hafta içinde şehirden ayrıldığı tespit edilmiştir (Talbot,2013).

İstatistik alanında büyük verinin dönüşümlere neden olacağı ileri sürülmektedir. Büyük veri setlerinin yönetimi ve analiz edilmesi, sigorta primlerinin hesaplanmasında aktüerya biliminin emek yoğun yaklaşımlarının kullanıldığı dönemlerden gelen uzun dönemli bir sorundur. Büyük veri bilimi bu köklerden gelmektedir. İstatistiğe dayalı geleneksel analitik yaklaşımlara ek olarak büyük veri, analitik algoritmaların düzenlenmesinde bilgisayar kaynakları ve yaklaşımlarından yararlanarak yeni teknikler getirmektedir. Tüm veritabanını işlemeyi sağlayan bilgi işlem alanındaki gelişmelerin örneklem kullanmayı gereksiz hale getirdiği ileri sürülmektedir (Erl vd.,2016:19).

### 3. İŞGÜCÜ PİYASALARININ İZLENMESİNDE BÜYÜK VERİ

Büyük veri bölgesel ve yerel işgücü piyasalarının izlenmesinde kullanılabilir. Bu sayede işgücü talebi ve işgücünde aranan niteliklerin hızlı ve ayrıntılı şekilde öngörülmesi mümkün olmaktadır. Büyük veri işgücü piyasası araştırmacılarının önceden incelenmesi zor olan konularda çalışmalarına katkıda bulunmaktadır. Örneğin Marinescu ve Rathelot (2014), ABD'li çevrimiçi iş sitesi Career Builder'den elde ettikleri verileri kullanarak, çalışanların iş başvurularında coğrafi etkilerin bir modelini belirlemeye çalışmışlardır. İş arayanların yaşadıkları yerden uzak işlere başvurmak istemedikleri ve coğrafi etkilerin işgücü piyasasında geçici işsizliği açıklamakta yetersiz olduğunu belirlemişlerdir. Büyük veri analizleri olmadan böyle bir araştırmanın yapılması zorlayıcı olacaktır.

Ayrıca büyük verinin geleneksel işgücü anketlerine göre üç tane avantajı olduğu ileri sürülmektedir. İlk olarak veri toplama maliyeti daha düşüktür. Veriler gözlemlenen durum bakımından sürekli güncellenmekte ve gerçek zamanlı analiz tekniklerinin kullanılmasını mümkün kılmaktadır. Son olarak ise önceden tanımlanmamış bir model veya sınıflandırmadan ziyade verilerden kaynaklandığı için sınıflandırma daha çeşitlidir. Örneğin geleneksel anketlerden farklı olarak, çevrimiçi açıklanan nitelikler önceden belirlenmiş sınıflandırmalara uymamakta, serbestçe ifade edilmektedir (Dusi vd.,2015:34)

İşsizliğin tahmin edilmesi ise büyük veri için en son gelişmedir. Bu durum 2008'den beri süren küresel ekonomik kriz nedeniyle kısa süreli işsizlik tahminlerine gerek duyulmasından kaynaklanmaktadır. Bu analizler, arama motorlarındaki çeşitli anahtar kelimeler ile aylık işsizlik oranlarının bileşimini belirleyen korelasyonların gözlenmesine dayalıdır. Gözlemlenen yapılar işsizlik oranlarını tahmin etmek için

kullanılmaktadır. Örneğin yüksek işsizlik zamanlarında internet kullanıcılarının hangi konularla ilgili olduğu incelenmektedir.

Uluslararası Çalışma Örgütü de (UÇÖ), işgücü istatistikleri bağlamında büyük veriye dikkat çekmektedir. UÇÖ 2014 Ekim ayında İsviçre’de yapılan ILO-Data2X toplantısında, büyük veri dahil olmak üzere teknolojinin daha iyi kullanımının önemi vurgulanmıştır. Ulusal istatistik bürolarından, uluslararası ve bölgesel örgütlerden ve sivil toplum temsilcilerinden oluşan bu grup özellikle enformel istihdam ve ücretsiz tarım işçiliği bağlamında cinsiyete dayalı veri açığına dikkat çekmiştir. Bu tartışmalar Birleşmiş Milletlerin 2015 sonrası kalkınma gündeminde veri devrimi ve kalkınma verilerin önemi şeklinde vurgulanmaktadır. Mobil telefon kayıtları, finansal modeller, çevrimiçi aktiviteler, algılama teknolojileri cinsiyet ile ilgili veri toplamasını arttıracak büyük veri çeşitleridir (Ferreira,2017:3).

#### 4. İŞGÜCÜ TALEBİNİN ÖLÇÜLMESİ AÇISINDAN BÜYÜK VERİ

İşgücü arzı ile işgücü talebinin eşleştirilmesi işgücü piyasasının en temel konusudur. Genellikle işgücü arzı yeterli ölçüde tanımlanmışken, işsizliğin doğası, iş arayan profilleri ve nitelikleri ile ilgili veri bulunurken, işgücü talebi yani firmaların boş işleri ve nitelik gereklilikleri konusunda yeterli bilgi elde etmek güçtür. İşgücü talebinin izlenebileceği internet, giderek daha fazla iş ilanı içeren önemli bir araç haline gelmiştir. İş ilanları yayınlayan çok sayıda portal bulunmaktadır. Bununla birlikte, internet üzerinden veri toplama tekniklerinin de büyük veri yaklaşımıyla geliştirildiği görülmektedir.

Literatürde işgücü talebinin ölçülmesinde internet veri kaynaklarının özellikle de iş arama siteleri ve gönüllü web anketlerinin incelendiği görülmektedir. Günümüzde web içerdiği büyük veri potansiyeli sayesinde coğrafi, mesleki ve nitelikler bağlamında işgücü piyasası dinamikleri hakkında önemli ipuçları sağlayabilmektedir. İş arama sitelerindeki ilanlar aracılığıyla yapılan metin analizlerinde işverenlerin talepleri belirlenebilmektedir. Ancak firmaların web siteleri aracılığıyla yayınladıkları açık işler, yani yapılandırılmamış metinlerden elde edilen enformasyonu işlemek için uygun bir yöntem izlenmelidir.

Dusi, Mercorio ve Mezzananza (2015) web sitelerinde işgücü talebi ile ilgili çok miktarda veri olduğunu belirterek, bu verilerin işgücü talebini yansıtacak kadar önemli olup olmadığını belirlenmesi gerektiğini ileri sürmektedir. Araştırmacılar Kuzey İtalya’da üç bölgede işgücü talebi ile ilgili web verilerini resmi verilerle karşılaştırmışlardır. Tüm çalışanlar için iş sözleşmesi başladığında, değiştirildiğinde ve sona erdirildiğinde bildirim yapılan zorunlu iletişim sistemi verileri ile web verileri karşılaştırılmaktadır. Veriler dört boyutta incelenmektedir: bölgesel, ekonomik sektör, sözleşme türü ve eğitim seviyesi. Web verileri ile resmi veriler, nicel özellik gösteren bölge ve sözleşme türü verileri açısından benzerlik göstermekteyken, nitel özellikler içeren eğitim seviyesi ile ekonomik sektör açısından farklılıklar tespit edilmiştir. Böylece işgücü talebi ile ilgili web verilerinin bazı yönlerden faydalı olabileceği tespit edilmiştir.

Ayrıca İtalya’da web üzerinden işgücü talebi enformasyonu toplayan İtalyan Digital İş İlanları İzleme prototipi WollyBi portalı oluşturulmuştur. WollyBi portalının hedeflediği gruplar, işe alma ve istihdam kurumları, kamu istihdam kurumları, işçi sendikaları ve mesleki eğitim aktörleridir. WollyBi portalının amacı, faaliyetlerinde bu aktörleri bölge, meslek ve nitelik boyutuyla derinlemesine işgücü piyasası talebi analizlerinde desteklemektir. WollyBi’nin veritabanı iki yıldan fazla süredir toplanan 1 milyon iş ilanından oluşmaktadır. WollyBi portalı verileri görselleştirmekte ve kullanıcılar nitelik, meslek ve iş sözleşmesi boyutlarında görsel olarak verileri analiz edebilmektedir (Dusi vd.,2015:39).

Çek Cumhuriyetinde, 1996 yılından beri açık işler izlenmektedir. Günümüzde ise açık işler için veri toplama süreci, tüm bölgelerdeki iş ofislerine kayıtlı açık işlerin ve çeşitli çevrimiçi iş arama sitelerindeki açık işlerin incelenmesi şeklinde gerçekleştirilmektedir. Bu sürecin sonunda, en çok talep edilen 25 meslek, talep edilen işler ile ilgili eğilimler, bölgesel analizler ve karşılaştırmalar ile belli sektörler ve eğitim alanları ile ilgili talep edilen eğitim seviyesi hakkında enformasyon sağlanmaktadır (Salavova,2015:50-51). Çek Cumhuriyeti örneğinde çoklu kaynaktan veri toplama faaliyetleri olduğunu görmekteyiz.

Literatürde iş arama sitelerindeki çeşitli meslekleri, cinsiyet ayrımcılığını, fazla kalifikasyonu, göçmenlerin durumunu inceleyen, bazılarında büyük veri tekniklerinin kullanıldığı araştırmalar da görülmektedir. Capiluppi ve Baravalle (2010), Monster.com sitesinden açık işleri indirmek için bir arama robotu geliştirmiş ve Birleşik Krallık’ta bilişim personeli için gereken niteliklerin içerik analizini yapmışlardır. Birleşik Krallık eğitim kurumlarının sundukları ile endüstrinin gereksinimleri arasında uyumsuzluk olduğunu belirlemişlerdir. Böylece büyük veri yöntemleri, işgücü piyasası ile eğitim kurumlarının ortak çalışması yönünde katkı sağlayabilmektedir.

Kuhn ve Shen (2013), Çin işgücü piyasasında işe alım sürecinde cinsiyet ayrımcılığını araştırmışlardır. Çin'in üçüncü büyük çevrimiçi iş portalında arama robotu ile veri toplanmıştır. Örneklem bir milyondan fazla iş ilanı ile firma verilerinin birleşmesinden meydana gelmektedir. Sonuç olarak cinsiyet ayrımcılığı olmadığı belirlenmiştir.

Kurekova ve arkadaşları (2013), Avrupa çapındaki EURES iş ilanları portalı verilerine dayalı olarak 3 küçük Avrupa ülkesinde (Çek Cumhuriyeti, Danimarka ve İrlanda) işverenlerin nitelik taleplerini karşılaştıran bir araştırma yapmışlardır. Bu ülkeler için nitelik isteklerinin çok farklı olduğunu ve genel bir gereklilikler seti olmadığı, yerel kurumların ve yapıların talebi şekillendirdiğini belirlemişlerdir. Araştırmacılar EURES verilerinin karşılaştırmalı araştırmalar için standart bir platform sunması ve Avrupa ülkelerinde yaygın kullanımı nedeniyle iyi bir kaynak olduğunu ileri sürmektedirler.

## 5. İŞSİZLİĞİN ÖLÇÜLMESİ AÇISINDAN BÜYÜK VERİ

2008 yazında Google firması Google Insights uygulamasının beta sürümünü kullanıma sokmuştur. Google Insights hizmeti ile, 2004 yılından itibaren dar veya geniş zaman ölçeklerinde, çeşitli anahtar kelimeler için arama sorguları, ülkeler ve bazı durumlarda bölgeler bazında karşılaştırılabilmektedir. Daha sonra hizmet Google Trends olarak adlandırılmıştır. Bir Google sorgusu bölgesel, zamansal ve anahtar kelime odaklı olabilmektedir. Sonuçlar, kullanıcılara ölçeklendirilmiş ve normalleştirilmiş olarak sunulmaktadır. (Askitas ve Zimmermann,2009:2) Böylece Google arama verilerine dayalı araştırmaların gündeme geldiğini görmekteyiz.

İşsizlik ile ilgili büyük veri araştırmalarına tıp alanındaki çalışmaların öncülük ettiği söylenebilir. Epidemiyoloji alanındaki araştırmalarda internet arama motoru verilerinin influenza gibi hastalıkların görülme sıklığını tahmin etmekte yardımcı olabileceği görülmüştür. Grip ile ilgili konuları aratan kişi sayısı ile grip belirtileri gösteren kişi sayısı arasında güçlü bir korelasyon olduğu belirlenmiştir. Araştırmacılar arama motoru verilerini geleneksel grip izleme sistemi ile karşılaştırdıklarında, arama sayılarında paralellik olduğunu tespit etmişlerdir (Ginsberg vd., 2009). Arama verilerine dayanarak hastalığın bölgelere yayılımını geleneksel hastalık izleme kurumlarından hızlı tespit etmek mümkündür.

Askitas ve Zimmermann (2009), Alman aylık işsizlik verileri ile belirledikleri Google anahtar arama kelimeleri arasındaki ilişkiyi incelemiştir. Öncelikle işsizlik bürosu, işsizlik oranı, kişisel danışman ve Almanya'da en popüler olan iş arama motorlarını anahtar kelime olarak belirlemişlerdir. Anahtar kelimeleri belirlerken bazı varsayımlardan yola çıkmışlardır. Örneğin işsizlik bürosu anahtar kelimesi, işsizlik bürosu ile iletişime geçmek isteyen kişilerin arama aktivitesi olarak görülmektedir. Kişisel danışmanlık ise işten çıkarılma korkusuna tepki veren yüksek nitelikli çalışanlarla ilişkili olması beklenen bir anahtar kelimedir. Anahtar kelimeler ile aylık işsizlik verileri arasında güçlü bir korelasyon olduğunu tesbit etmişlerdir.

Vicente ve arkadaşları (2015), arama motoru verilerine dayalı araştırmaların genelde orta veya düşük işsizlik oranı olan ülkelerde yapıldığını belirterek, işsizlik oranının yüksek olduğu İspanya için bir araştırma yapmışlardır. Zaman serisi yaklaşımının kullanıldığı araştırmada, iş duyarlılığı endeksi (istihdam güven göstergesi) ve internet arama yoğunluğu (Google Trends aracılığıyla) değişkenleri analiz edilmiştir. İstihdam güven endeksi, aylık yayınlanan ve kamu istihdam hizmetlerine kayıtlı iş taleplerini gösteren İspanya İstihdam ve Sosyal Güvenlik Bakanlığı verilerinden alınmaktadır. Diğer değişken ise google trends verilerinden "iş teklifi" anahtar kelimesiyle oluşturulmuştur. 2004-2013 yıllarını kapsayan araştırma sonucunda İspanya'da işsizliğin kapsamlı bir değerlendirilmesi yapılmış ve ekonomik krizler bağlamında işsizlik istatistiklerinin tahmin edilmesi sağlanmıştır. Araştırma sonuçları internet verilerinin ekonomik değişkenleri izlemek ve tahmin etmekte yararlı olduğunu göstermektedir.

Barreira, Godinho ve Melo (2013), dört farklı ülkeden sosyal ve ekonomik anlamda değeri olan iki alanı (işsizlik oranı ve araba satışları) basit otoregresif modeller aracılığıyla internet arama verilerini kullanarak incelemiştir. İşsizlik oranları açısından Google Trends verilerinin dört ülkeden üçünde tahminleri iyileştirdiğini tespit etmişlerdir: Portekiz, Fransa ve İtalya. Bununla birlikte, farklı örneklem dışı dönemler dikkate alındığında bu verilerin öngörü yeteneklerinde önemli farklılıklar görülmektedir.

Google Trends uygulamasından elde edilen verilerle işsizliğin öngörülenmesinin yanında cep telefonu verilerine dayalı araştırmalar da yapılmaktadır. Toole ve arkadaşları (2015), cep telefonlarından elde edilen veriler ile ekonomik şokların ve sonuçlarının izlenmesinin mümkün olabileceğini ileri sürmektedir. Cep telefonlarından elde edilen çağrı kayıtlarını kullanarak, toplu işten çıkarmaları, bundan etkilenen bireyleri belirlemek ve genel işsizlikteki değişimleri tahmin edebilen orijinal bir yöntem geliştirmişlerdir. Kapanan bir fabrikayı kullanarak, toplu işten çıkarma zamanını ve işten çıkarmanın boyutlarını belirlemişlerdir. Fabrika kapanmasından sonra bireylerin çağrı davranışlarının nasıl etkilendiğini gösteren bir model geliştirmişlerdir.

Model işten çıkarılan bireylerin hareketliliklerinde ve sosyal davranışlarında önemli azalmalar olduğunu göstermektedir. Mikro düzeyde belirlenen bu özellikleri kullanarak, çağrı davranışlarındaki değişimlerin, bölgesel düzeyde toplandığında, makro işsizlik verilerini öngörümlemeyi geliştirebilecektir. Bu yöntem ve sonuçlar mikro ekonomik davranışları ölçmek için yeni bir veri kaynağına dikkat çekmektedir.

## 6. İŞGÜCÜ PİYASASINDA BÜYÜK VERİ İLE İLGİLİ SORUNLAR

İşgücü piyasasında büyük veri uygulamaları konusunda yaşanan önemli zorluklardan bir tanesi örneklem ve seçim konusundadır. Büyük veri kaynakları genellikle sınırları tam bilinmeyen örneklemelerden oluşmaktadır. Hanehalkı işgücü anketlerinin tersine, çevrimiçi işgücü aracılığı faaliyetlerinden toplanan veriler için örneklem çerçevesi belirgin olmayabilmektedir. Çevrimiçi işgücü piyasalarına, bilgi alışverişine, iş ilanlarına katılan kullanıcılar için farklı güdülerin bulunması, örneklem içinde kimin olacağı, neden olacağı, ne zaman katkıda bulunacağı ve sağlanan enformasyonun doğruluğunu etkileyen faktörlerdir (Baruffini,2015:264).

Ancak örneklem kullanılan veri toplama tekniğine bağlı olarak çeşitli düzeylerde kontrol edilebilir. Örneğin hedefli e-posta ve davet edilen kişilerden alınan cevaplar sayesinde, planlanmış ve bilinen bireylerden oluşan bir örneklemde veri üretilebilir. Araştırma sonuçlarına göre, e-posta çeşitli özendiriciler ve takip ile birlikte kullanıldığında, geri dönüş oranları geleneksel yöntemlerle karşılaştırılabilecek seviyededir. Fakat bu stratejinin dezavantajı, katılımcının anonimliğinin ortadan kalkmasıdır. Bir internet sayfasının katılım için kullanılması ise, daha fazla anonimlik sağlamakta, fakat örneklem açısından sorun yaratmaktadır (Baruffini,2015:265).

Büyük veri, düşünüldüğünden daha az sağlam olabilir. Örneğin google arama verileri kullanılarak yapılan grip araştırmalarında olduğu gibi, belli bir zamandaki veri yapısı diğer bir zamandakine uymayabilir. Web sitelerindeki verilere farklı şekillerde farklı amaçlarla doğru olmayan sonuçlar verecek şekilde müdahale edilebilir. Buna ek olarak, veri toplamada sık kullanılan arama motorları endekslerinde müdahaleler yapılması (spamdexing, google bombing) mümkündür (Lake ve Drake,2014:7).

## 7. TÜRKİYE'DE İŞSİZLİĞİN ÖNGÖRÜLMESİNDE GOOGLE TRENDS VERİLERİ İLE BİR ARAŞTIRMA

İşsizliğin tahmin edilmesinde Google arama verilerinin kullanımı büyük veri için iyi bir örnek olarak görülmektedir. Yukarıda değinildiği üzere, Google arama verilerinin işsizlik oranlarının tahmin edilmesinde kullanılabileceğini gösteren birçok araştırma bulunmaktadır. Ülkemizde bu tür araştırmalara çok az rastlanmaktadır.

Chadwick ve Şengül (2012), Google tarama sonuçları verilerinin Türkiye'de üç aylık bir gecikme ile yayınlanan aylık tarım dışı işsizlik dönem içi tahminini geliştirip geliştirmediğini araştırmıştır. Bunun için yapılan analizde lineer tahmin modelleri ve Bayesgil Model Ortalaması yöntemi ve Ocak 2005 ve Kasım 2011 arası verisi kullanılmıştır. Sonuçlar, Google tarama sonuçları verilerinin hem örneklem içinde hem de örneklem dışında tarım dışı işsizlik oranı dönem içi tahminlerini iyileştirdiğini göstermektedir.

### 7.1. Araştırmanın Amacı

Araştırmamızda 15 ve üstü yaş ile 15-24 yaş grubunun işsizlik oranları ve 4 Google arama terimi arasındaki ilişki incelenmiştir. Büyük verinin işgücü piyasasında kalitesini değerlendirmek için başka kaynaktan elde edilen verilerle karşılaştırılması önerilmektedir. Bu nedenle Türkiye İstatistik Kurumunun (TÜİK) işsizlik verileri kullanılmıştır. Çalışmanın amacı, Google arama verilerinin işsizliğin tahmininde kullanılıp kullanılamayacağını belirlemektir.

### 7.2. Veri Seti

Araştırmada Google'da işsizlik ile ilgili yapılan taramalar ile işsizlik oranı tahmin edilmeye çalışılmıştır. Bu amaçla mevsimsel olarak düzeltilmiş\* aylık işsizlik oranları kullanılmıştır. Araştırmada işsizlik oranı, 15-24 yaş ve 15 ve üstü yaş için iki ayrı kategoride incelenmiştir. Veri seti 2005:M1-2017:M7 dönemini kapsamaktadır. İşsizlik verileri TÜİK resmi web sitesinden elde edilmiştir.

İşsizliğin tahmin edildiği modele, Google web sitesi aracılığıyla yapılan iş arama ile ilgili ifadeleri özetleyen aylık Google Index değişkeni dışsal değişken olarak dahil edilmiştir. Bu kapsamda Google web sitesinde "iş arama", "iş bulma", "iş ilanı" ve "eleman arama" anahtar kelimeleri ile yapılan sorgular dikkate alınmıştır. Söz konusu dört anahtar kelimeyi dikkate almamızın nedeni, farklı iş arama ile ilgili aramalar arasında bu

\* Verilerin mevsimsel olarak düzeltmek amacıyla Tramo-Seats yönteminden yararlanılmıştır.

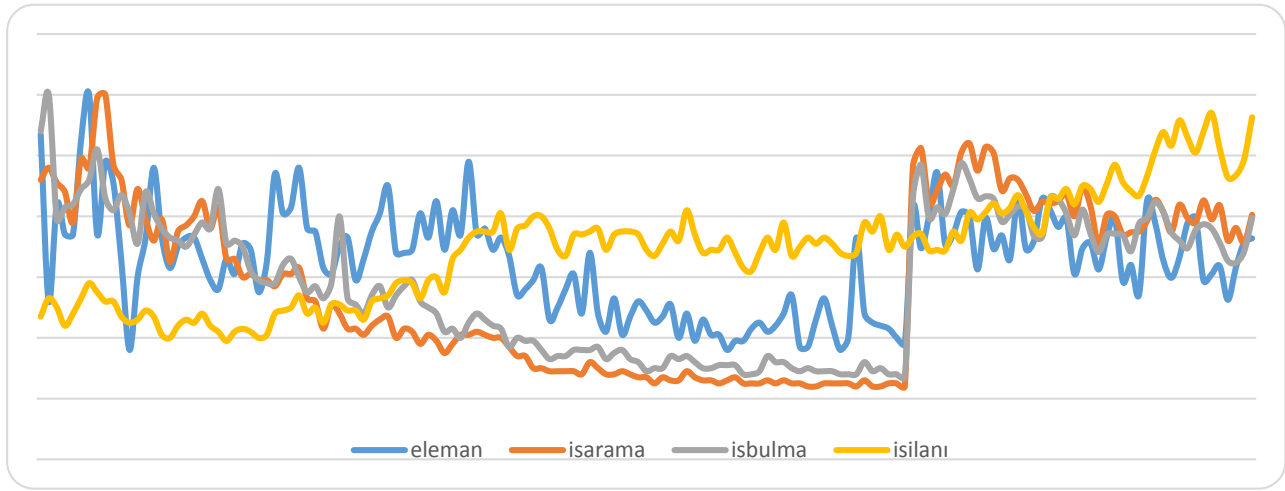
ifadelerin en yüksek orana sahip olmalarıdır. Bir diğer neden ise, iş arayanların büyük çoğunluğunun bu kelimeleri yaygın olarak kullanmalarıdır. Google Index'e ilişkin veriler en yüksek değer olan 100 ile normalize edilerek ele alınmıştır. Tablo 1'de araştırmada kullanılan değişkenlere ilişkin açıklamalar yer almaktadır:

Tablo 1. Kullanılan Değişkenlere İlişkin Açıklamalar

Değişkenler	Açıklamaları
issizlik	15 yaş ve üstü işsizlik oranları
issizlikg	15-24 yaş arasındaki bireyler için işsizlik oranları
eleman	"Eleman arıyor" ifadesine ilişkin Google Index değeri
isarama	"İş arıyor" ifadesine ilişkin Google Index değeri
isbulma	"İş bulma" ifadesine ilişkin Google Index değeri
isilanı	"İş ilanı" ifadesine ilişkin Google Index değeri

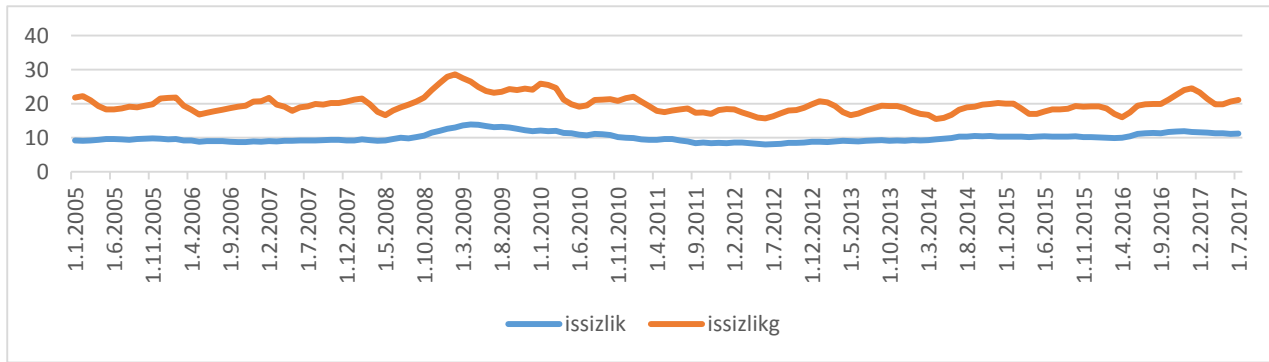
Şekil 1'de iş arama", "iş bulma", "iş ilanı" ve "eleman arama" sorguları için aylık ortalama değerlerine ilişkin grafikler yer almaktadır:

Şekil 1. İşsizlik İle İlgili Google Index Değerleri



Şekil 2'de ise 15-24 yaş ve 15 yaş üstü için işsizlik oranlarına ilişkin grafikler sergilenmektedir:

Şekil 2. 15-24 Yaş ve 15 Yaş Üstü İşsizlik Oranları



Tablo 2'de değişkenlere ilişkin özet istatistikler yer almaktadır:

Tablo 2. Değişkenlere İlişkin Özet İstatistikler

	issizlik	issizlikg	eleman	isarama	isbulma	isilanı
<b>Ortalama</b>	10.02450	19.86234	44.81026	37.61358	38.67506	49.07660
<b>Medyan</b>	9.600000	19.25859	46.00000	37.00000	39.00000	50.00000
<b>Maksimum</b>	13.90000	27.80191	100.0000	100.0000	100.0000	94.00000
<b>Minimum</b>	8.000000	16.17826	16.00000	4.000000	7.000000	19.00000
<b>Standard Sapma</b>	1.310571	2.241548	16.79306	27.48891	22.84613	17.75833
<b>Çarpıklık</b>	0.991410	1.299329	0.287026	0.199111	0.111951	0.298113
<b>Basıklık</b>	3.409983	4.756757	2.899830	1.608746	1.811961	2.597246

Tablo 2 incelendiğinde, 2005M1-2017M7 dönemi için 15 yaş üstü ve 15-24 yaş arası işsizlik oranları için ortalamalar sırasıyla %10.02 ve %19.86'dır. Söz konusu dönemde 15 yaş üstü ve 15-24 yaş arası işsizlik

oranlarına ilişkin en yüksek değerler ise %13.9 ve %27'dir. Ayrıca, genç işsizliğe (15-24 yaş arası işsizlik) ilişkin değişkenliğin 15 yaş üstü işsizliğe göre daha yüksek olduğu ifade edilebilir. "iş arama", "iş bulma", "iş ilanı" ve "eleman arama" ifadelerine ilişkin Google Index değerleri karşılaştırıldığında, yapılan en yüksek aramanın "iş ilanı" ifadesine ait olduğu görülmektedir. Söz konusu ifadeler değişkenlik açısından karşılaştırıldığında ise, en yüksek değişkenliğin "iş arama" ifadesinde olduğu tespit edilmiştir. Tüm değişkenler sağa çarpık bir dağılıma sahiptir ve basıklık değerleri yüksektir. Bu durum serilerin normal dağılmadığını ifade etmektedir.

### 7.3. Ampirik Bulgular

Çalışmada işsizlik oranını tahmin etmek amacıyla ARIMA modellerinden yararlanılmıştır. ARIMA modellerinin uygulanabilmesi için serilerin durağan olmaları gerekmektedir. Bu nedenle, uygun ARIMA modelleri belirlenmeden önce ilk olarak serilerin durağanlıkları ADF ve Phillips-Perron birim kök testleri ile araştırılmıştır. Sonuçlar Tablo 3'de yer almaktadır:

Tablo 3. ADF ve Phillips-Perron Birim Kök Test Sonuçları

	ADF		Phillips-Perron	
	Sabit terimli	Sabit terimli ve trendli	Sabit terimli	Sabit terimli ve trendli
İşsizlik	-1.931	-1.967	-1.833	-1.881
Δİşsizlik	-3.632***	-3.631***	-8.730***	-8.705***
İşsizlikg	-2.361	-2.336	-1.972	-1.939
Δİşsizlikg	-4.076***	-4.075***	-8.921***	-8.900***
Eleman	-2.330	-2.279	-5.533***	-5.566***
Δeleman	-11.231***	-11.208***	-26.239***	-25.869***
İsarama	-1.757	-1.522	-1.600	-1.609
Δisarama	-13.973***	-14.033***	-13.983***	-14.046***
İsbulma	-2.118	-2.035	-1.941	-1.815
Δisbulma	-13.167***	-8.709***	-13.536***	-13.884***
İsilanı	-0.915	-2.697	-1.098	-3.450**
Δisilanı	-2.904**	-2.863	-21.831***	-21.768***
Kritik değerler	%1 -3.477 %5 -2.882 %10 -2.577	%1 -4.024 %5 -3.442 %10 -3.145	%1 -3.474 %5 -2.880 %10 -2.577	%1 -4.020 %5 -3.440 %10 -3.144

Not: \*\*\*, \*\*, \* sırasıyla %1, %5 ve %10 önem seviyelerinde istatistiki anlamlılıkları ifade etmektedir. Δ değişkenlerin birinci farklarının alındığını göstermektedir.

Tablo 3'de ADF birim kök testi dikkate alındığında, hem sabit terimli hem de sabit terim ve trendli birim kök test sonuçlarına göre tüm değişkenlerin düzey değerlerine birim kök içerdiği görülmektedir. Buna karşın, Phillips-Perron birim kök testine göre eleman ve işilanı değişkenleri 0.05 önem seviyesinde düzey değerlerinde durağandır. ADF ve Phillips-Perron birim kök testlerinin birbirine zıt sonuçlar vermesi seride yapısal kırılma olduğunu göstermektedir. Aynı zamanda, ele alınan dönemin 2008 küresel finansal krizini kapsamı ve serilerin grafikleri, söz konusu değişkenlerde yapısal kırılmanın varlığına işaret etmektedir. Bu nedenle çalışmada tek yapısal kırılmaya izin veren birim kök testi olan Zivot Andrews birim kök testi uygulanmıştır. Sonuçlar Tablo 4'de yer almaktadır:

Tablo 4. Zivot Andrews Birim Kök Test Sonuçları

	Sabit terimli	Kırılma tarihi	Sabit terimli ve trendli	Kırılma tarihi
İşsizlik	-4.015	2010M10	-4.220	2010M9
Δİşsizlik	-5.528***	2009M5	-6.135***	2009M5
İşsizlikg	-3.476	2010M9	-3.481	2010M4
Δİşsizlikg	-6.363***	2009M5	-6.846***	2009M5
Eleman	-5.855***	2014M1	-5.519**	2014M1
İsarama	-14.458***	2014M1	-16.710***	2014M1
İsbulma	-13.955***	2014M1	-15.524***	2014M1
İsilanı	-4.835*	2008M6	-5.151**	2007M6
Kritik değerler	%1 -5.34 %5 -4.93 %10 -4.58		%1 -5.57 %5 -5.08 %10 -4.82	

Not: \*\*\*, \*\*, \* sırasıyla %1, %5 ve %10 önem seviyelerinde istatistiki anlamlılıkları ifade etmektedir. Δ değişkenlerin birinci farklarının alındığını göstermektedir.



Tablo 4’de Zivot Andrews birim kök testi incelendiğinde, hem sabit terimli hem de sabit terim ve trendli test sonuçlarına göre 15-24 yaş ve 15 yaş üstü işsizlik oranlarının birinci farklarında, “iş arama”, “iş bulma”, “iş ilanı” ve “eleman arama” ifadelerine ilişkin Google Index değerlerinin ise düzey değerlerinde durağan oldukları görülmektedir. Ayrıca, hem 15-24 yaş hem de 15 yaş üstü işsizlik oranlarındaki kırılma tarihleri, kriz dönemlerine denk gelmektedir. Benzer şekilde “iş ilanı” ifadesi ile Google’da yapılan aramaların kırılma tarihi de yine krizin etkisini yansıtmaktadır. Buna karşın “iş arama”, “iş bulma” ve “eleman arama” ifadeleri ile Google’da yapılan aramalarda meydana gelen kırılmalar ise kriz sonrası dönemi göstermektedir.

Çalışmada Google Index’i ile işsizlik öngörüsünde bulunmak amacıyla ARIMA modellerinden yararlanılmıştır. ARIMA modelleri hem serinin geçmiş dönemdeki gözlem değerlerinin hem de geçmiş dönem şoklarının söz konusu seri üzerindeki etkilerini ortaya koymaktadır. Uygun ARIMA modelinin seçilebilmesi için seriye ilişkin otokorelasyon ve kısmi otokorelasyon fonksiyonları ve bilgi kriterleri dikkate alınmıştır. Google Index ile 15-24 yaş ve 15 yaş üstü işsizlik oranları tahmin etmek için en uygun modellerin sırasıyla ARIMA(2,1,2) ve ARIMA(1,1,2) modeli olduğu görülmektedir. Kurulan modeller aşağıdaki gibi ifade edilebilir:

$$\Delta \text{issizlik}_t = \beta_0 + \beta_1 \Delta \text{issizlik}_{t-1} + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \varepsilon_t$$

$$\Delta \text{issizlik}_t = \beta_0 + \beta_1 \Delta \text{issizlik}_{t-1} + \beta_2 \text{Google Index}_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \varepsilon_t$$

$$\Delta \text{issizlik}_g_t = \beta_0 + \beta_1 \Delta \text{issizlik}_{t-1} + \beta_2 \Delta \text{issizlik}_{t-2} + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \varepsilon_t$$

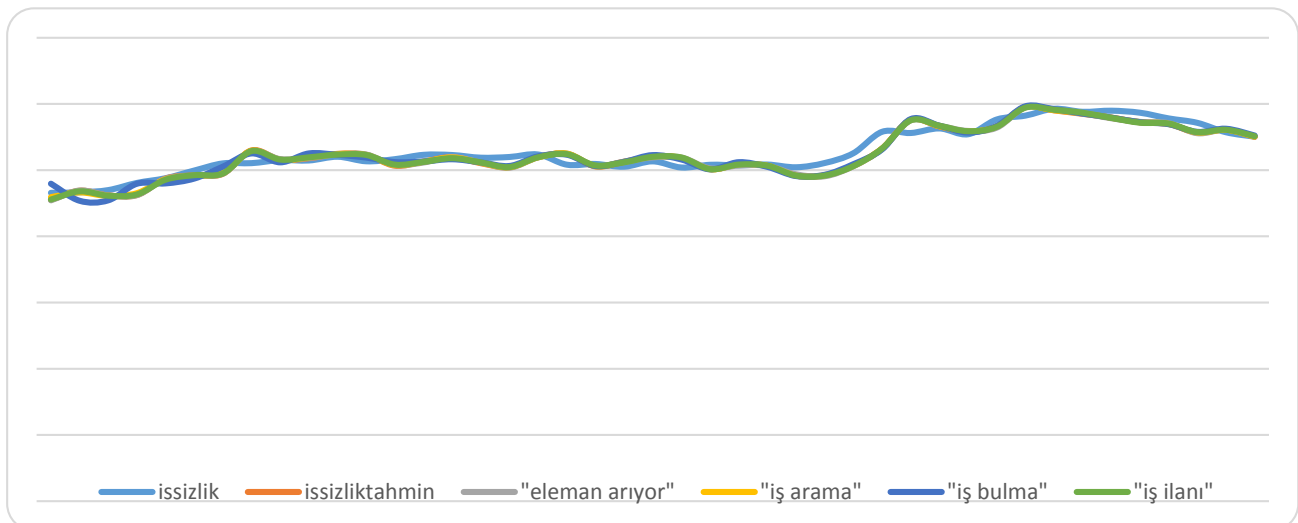
$$\Delta \text{issizlik}_g_t = \beta_0 + \beta_1 \Delta \text{issizlik}_{t-1} + \beta_2 \Delta \text{issizlik}_{t-2} + \beta_3 \text{Google Index}_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \varepsilon_t$$

Modellerde Google Index ile ifade edilen terim “iş arama”, “iş bulma”, “iş ilanı” ve “eleman arama” ifadelerini,  $\varepsilon_t$  şokları yansıtan hata terimini,  $\beta$  ve  $\alpha$  ise sırasıyla AR ve MA katsayılarını göstermektedir. Tablo 5’de oluşturulan ARIMA modellerine ilişkin istatistiksel tahmin sonuçları yer almaktadır. Tablo 5’e göre gerek 15-24 yaş gerekse 15 yaş üstü işsizlik oranlarının tahmininde “iş arama” ve “iş bulma” ifadelerine ilişkin Google Index değerlerinin etkili olduğu ifade edilebilir. Ayrıca, Her iki işsizlik oranı için de iki dönem (iki ay) önceki işsizlik rakamları cari dönem işsizlik oranının artmasına neden olmaktadır. Benzer şekilde, geçmiş dönemdeki şoklar da işsizlik oranlarını arttıran bir faktör olarak karşımıza çıkmaktadır.

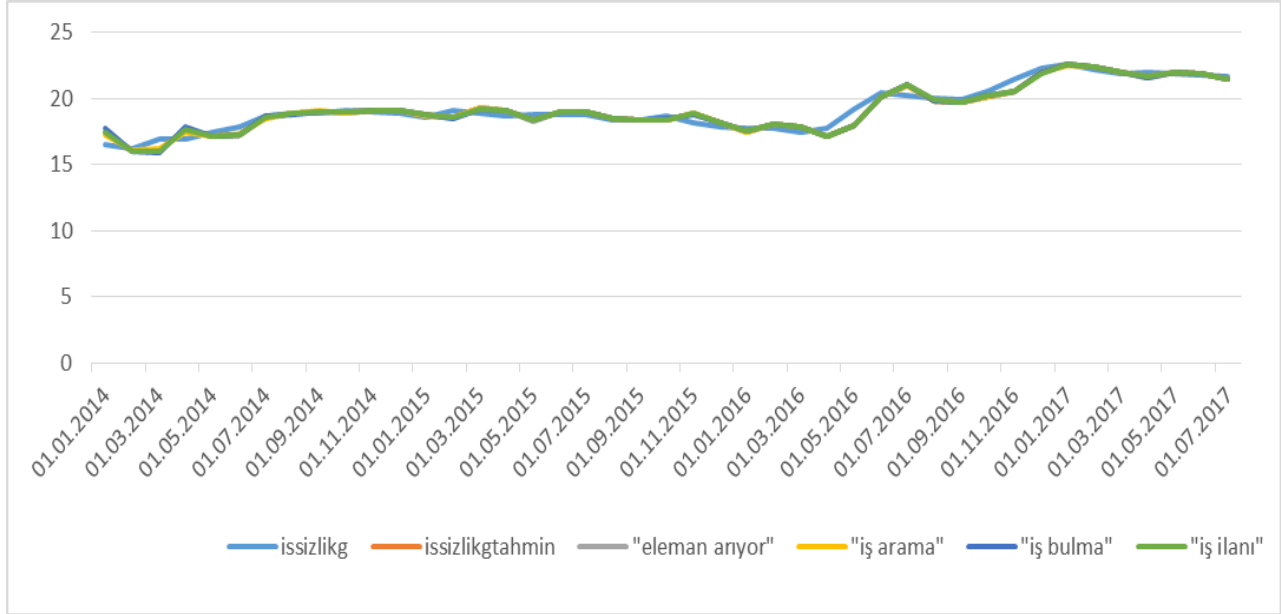
Yapılan analizlerle modellerin uygunluğu tespit edildikten sonra, her iki işsizlik serisi için oluşturulan ARIMA(1,1,2) ve ARIMA(2,1,2) modelleri ile 2014M1-2017M7 dönemi için işsizlik oranı tahminleri gerçekleştirilmiştir. Bu amaçla modeller 2005M1-2013M12 dönemini kapsayacak şekilde kurulmuştur. Böylelikle, gerçekleşen ve tahmin edilen değerler arasındaki farklılık dikkate alınarak tahmin edilen parametrelerin geçerliliği tespit edilmiştir.

Şekil 3’de “iş arama”, “iş bulma”, “iş ilanı” ve “eleman arama” ifadelerine ilişkin Google Index değerleri ile yapılan işsizlik tahmin değerlerine ilişkin grafikler yer almaktadır. Şekil 3’e göre, gerçekleşen ve tahmin edilen işsizlik değerlerinin birbirine yakın olduğu görülmektedir.

Şekil 3. İşsizlik Öngörüsüne İlişkin Grafikler  
a. 15 Yaş Üstü İşsizlik Oranları



## b. 15-24 Yaş İşsizlik Oranı



Gerçekleşen ve tahminlenen işsizlik değerlerinin birbirine benzerliğinin net bir şekilde görülebilmesi amacıyla 2017 yılının ilk 7 ayı için söz konusu değerler Tablo 6'da gösterilmektedir.

Tablo 6. Gerçekleşen ve Öngörülen İşsizlik Değerleri

15 Yaş Üstü İşsizlik Oranları						
	Gerçekleşen	Öngörülen				
t	issizlik	issizliktahmin	"eleman arıyor"	"iş arama"	"iş bulma"	"iş ilanı"
01.01.2017	11.76409	11.70309869	11.72577518	11.71411749	11.70760674	11.73703
01.02.2017	11.79695	11.59058423	11.56871399	11.58256124	11.58669883	11.58003
01.03.2017	11.73549	11.44509365	11.45961513	11.4502678	11.44959008	11.4357
01.04.2017	11.564	11.39501424	11.4186222	11.39476713	11.38769533	11.4086
01.05.2017	11.43017	11.13453407	11.11794089	11.13811661	11.15615073	11.15053
01.06.2017	11.14502	11.22234631	11.22379186	11.22036746	11.25098481	11.22204
01.07.2017	11.01127	11.0060222	11.02266935	11.01351623	11.04572403	11.0268
15-24 Yaş İşsizlik Oranları						
	Gerçekleşen	Öngörülen				
t	issizlikg	issizliktahmin	"eleman arıyor"	"iş arama"	"iş bulma"	"iş ilanı"
01.01.2017	22.61576	22.55617392	22.5735881	22.50831	22.570387	22.597
01.02.2017	22.18483	22.38467853	22.37678387	22.404059	22.378268	22.405
01.03.2017	21.92751	21.9859891	21.97787237	21.990559	21.975793	21.9438
01.04.2017	21.96233	21.62519015	21.65421197	21.614177	21.621951	21.6338
01.05.2017	21.88462	21.96785501	21.94389367	21.974022	21.970683	21.9944
01.06.2017	21.75603	21.84722074	21.83242984	21.84876	21.864652	21.8447
01.07.2017	21.63019	21.46017752	21.48343719	21.431786	21.490291	21.503

Gerçekleşen ve tahminlenen değerler arasındaki farkın birbirine yakın olması yapılan tahminlerin güvenilirliğini göstermektedir. Bu durum öngörü performansı olarak da adlandırılmaktadır. Çalışmada öngörü performansının değerlendirilmesi amacıyla kullanılan RMSE, MAE, MAPE ve Theil eşitsizlik katsayısı ölçütlerinden yararlanılmıştır. Sonuçlar Tablo 7'de gösterilmektedir. Tablo 7 incelendiğinde, kurulan modellerin öngörü performanslarının yüksek olduğu ifade edilebilir. Ayrıca, Google Index değerleri karşılaştırıldığında her iki işsizlik değerinin tahmin edilmesinde "iş arama" ve "iş bulma" ifadeli Google Index değerlerinin en iyi öngörü performansı sağladığı görülmektedir.

Tablo 7. Öngörü Performanslarına İlişkin Ölçütler

15 Yaş Üstü İşsizlik Oranı					
	Google Index'in Olmadığı Model	“eleman arıyor” ifadeli Google Index	“iş arama” ifadeli Google Index	“iş bulma” ifadeli Google Index	“iş ilanı” ifadeli Google Index
RMSE	0.212083	0.211945	0.210039	0.203099	0.210430
MAE	0.165719	0.167096	0.163850	0.161265	0.164743
MAPE	1.648386	1.660450	1.629771	1.601151	1.638076
Theil Eşitsizlik Katsayısı	0.010453	0.010446	0.010352	0.010009	0.010371
15-24 Yaş Arası İşsizlik Oranı					
	Google Index'in Olmadığı Model	“eleman arıyor” ifadeli Google Index	“iş arama” ifadeli Google Index	“iş bulma” ifadeli Google Index	“iş ilanı” ifadeli Google Index
RMSE	0.475523	0.484219	0.478580	0.492515	0.483032
MAE	0.372181	0.379358	0.378168	0.381341	0.378520
MAPE	1.873179	1.909731	1.899992	1.920880	1.904220
Theil Eşitsizlik Katsayısı	0.011902	0.012116	0.011975	0.012324	0.012086

## 8. SONUÇ

Büyük veri uygulamalarının toplumda çok yönlü etkilerinin olacağı anlaşılmaktadır. Büyük veri sahip olduğu hız, doğruluk, çeşitlilik ve büyüklük gibi özellikleri sayesinde yeni ekonominin önemli bir bileşeni olarak görülmektedir. Ayrıca karar alma süreçleri üzerindeki etkileri ve bu anlamda potansiyelinin yüksekliği vurgulanmaktadır. İşgücü piyasasının değerlendirilmesi açısından da etkileri olacağı yazında açıklanmaktadır.

İşgücü piyasalarının izlenmesinde yeni yöntemlerin ortaya çıkacağı ve daha önce incelenmeyen yönlerin analiz edilebileceğini gösteren araştırmalar bulunmaktadır. İlk olarak büyük veri, işletmelerin işgücü talebi hakkında farklı bilgiler edinilmesini sağlayabilmektedir. İşgücü talebi, işgücü piyasası analizlerinde genellikle yeterince bilgi toplanamayan bir boyut olarak karşımıza çıkmaktadır. İş ilanlarının internet ortamına taşınması, bu ilanları toplayabilecek büyük veri yazılımları geliştirilmesi sonucunda işletmelerin nitelik talepleri, cinsiyet ayrımcılığı yapıp yapılmadığı gibi konular araştırılabilmektedir.

İkinci olarak, işsizliğin izlenmesinde geleneksel veri kaynaklarına ek kaynakların ortaya çıktığı görülmektedir. Özellikle Google Trends verilerine dayalı olarak işsizliği inceleyen çok sayıda araştırma olduğu görülmektedir. Google Trends verilerinin işsizlik tahminlerini iyileştirdiği ve geleneksel işsizlik izleme yöntemlerine ek olarak kullanılabilmesi tespit edilmiştir. Bunun yanında cep telefonu verilerinin işsizlik ile ilgili öngörülerde bulunmak için kullanılabilmesini gösteren araştırmalar da bulunmaktadır. Ayrıca Twitter, Facebook gibi sosyal medya uygulamalarının da işgücü piyasası boyutunda analiz potansiyeli olduğu ileri sürülmektedir.

Google arama verilerinin ülkemizde işsizliğin tahmininde kullanılıp kullanılmayacağını belirlemek üzere bir araştırma yapılmıştır. Araştırmamızda 15 ve üstü yaş ile 15-24 yaş grubunun işsizlik oranları ve 4 Google arama terimi ekonometrik modellerle incelenmiştir. İşsizlik oranını tahmin etmek için ARIMA modellerinden yararlanılmıştır. Bulgulara göre gerçekleşen ve tahmin edilen işsizlik değerlerinin birbirine yakın olduğu görülmektedir. Özellikle “iş arama” ve “iş bulma” arama kelimelerinin en iyi performansı sağladığı görülmektedir. Kısaca büyük veri kaynaklarından biri olan Google arama motorunun işsizliğin tahminlenmesinde katkıda bulunabileceği anlaşılmaktadır.

Google tabanlı araştırmalarda örneklemin belirsiz olması araştırmanın kısıtıdır. Bu anlamda, Google verilerinin işsizliğin öngörülmesinde resmi istatistiklere yardımcı olarak düşünülmesi önerilebilir. Başka bir deyişle, resmi istatistiklerin çoklu veri kaynaklarından (Google, cep telefonu çağrıları, sosyal medya vb.) sağlanması yönünde uygulamalar geliştirilebilir. Büyük veri yaklaşımı ayrımcılık, yoksulluk, barınma gibi sosyal politikanın çeşitli boyutlarının araştırılmasında kullanılacak imkânlarda sunmaktadır.

## KAYNAKLAR

Askitaş, N. & Zimmermann, K. (2009). “Google Econometrics and Unemployment Forecasting.” IZA Discussion Papers No. 4201. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1465341](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1465341) (erişim 05.10.2017).

- Barreira N., Godinho P., Melo P. (2013). Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends. *NETNOMICS: Economic Research and Electronic Networking*. Volume 14, Issue 3, pp 129–165.
- Baruffini M., (2015). Assessing Data Collection and Data Quality for Labour Market Monitoring. (Ed.) Larsen C., Rand S., Schmid A., Mezzanzanica M., Dusi S. *Big Data and the Complexity of Labour Market Policies: New Approaches in Regional and Local Labour Market Monitoring for Reducing Skills Mismatches*. München: Rainer Hampp Verlag.
- Capiluppi, A., Baravalle A. (2010). “Matching Demand and Offer in On-Line Provision: A Longitudinal Study of Monster.com.” In . Romania. <http://roar.uel.ac.uk/995/> (erişim 17.10.2017).
- Chadwick M. G., Şengül G. (2012). Nowcasting Unemployment Rate in Turkey: Let's Ask Google. Central Bank of the Republic of Turkey Working Paper No: 12/18.
- Chen M., Mao S., Zhang Y., Leung Victor C.M. (2014). *Big Data Related Technologies, Challenges and Future Prospects*. Springer.
- DeRoos, D. (2012). What is Big Data and how does it fit into an Information Integration Strategy. In *Information Integration & Governance Forum 2012*. Phoenix: IBM Corporation.
- Dorschel W., Dorschel J. (2015). Einführung. (Ed.) Dorschel Joachim, *Praxishandbuch Big Data Wirtschaft – Recht – Technik*. Wiesbaden:Springer Gabler.
- Dusi S., Mercurio F., Mezzanzanica M., (2015). Big Data Meets Web Job Vacancies: Trends, Challenges and Development Directions. (Ed.) Larsen C., Rand S., Schmid A., Mezzanzanica M., Dusi S. *Big Data and the Complexity of Labour Market Policies: New Approaches in Regional and Local Labour Market Monitoring for Reducing Skills Mismatches*. München: Rainer Hampp Verlag.
- Erl T., Khattak W., Buhler P. (2016). *Big Data Fundamentals Concepts, Drivers & Techniques*. Prentice Hall.
- Ferreira, J. (2017). International Labour Organization. (Ed.) L.A. Schintler, C.L. McNeely, *Encyclopedia of Big Data*, Springer International Publishing AG.
- Ginsberg, J., Mohebbi, M.H., Patel, R.J., Brammer, L., Smolinski, M.S. Brilliant, L. (2009). “Detecting influenza epidemics using search engine query data”, *Nature*, 457: 1012-1014.
- King, S. (2013). *Big Data Potential und Barrieren der Nutzung im Unternehmenskontext*. München: Springer VS.
- Klous S., Wielaard N. (2016). *We are Big Data The Future of the Information Society*. Amsterdam: Atlantis Press.
- Kuhn, P.& Kailing S. (2013). “Gender Discrimination in Job Ads: Evidence from China.” *The Quarterly Journal of Economics* 128 (1): 287–336.
- Kureková, L., Beblavý M., Haita C., Thum A. (2013). “Demand for Low-Skilled Workers across Europe: Between Formal Qualifications and Non-Cognitive Skills.” *NEUJOBS Working Paper No. 4.3.3*.
- Lake P. & Drake R. (2014). *Information Systems Management in the Big Data Era*. Sheffield:Springer.
- Marinescu, I. & Rathelot, R. (2014). Mismatch unemployment and the geography of job search. Working Paper 22672 <http://www.nber.org/papers/w22672> (erişim 10.09.2017).
- McKinsey Global Institute. (2011). *Big data: the next frontier for innovation, competition and productivity*. McKinsey Global Institute, New York.
- Page, C., Campbell, R., Coggeshall, S., Gillespie, E., Johnson, R., Olson, M. & Perkins, P. (2012). *Demystifying Big Data: A Practical Guide To Transforming The Business of Government* Microsoft. Washington, DC.
- Talbot, D. (2013). Big Data from cheap phones. <http://www.technologyreview.com/featuredstory/513721/big-data-from-cheap-phones/> (erişim 10.09.2017).
- Toole J.L., Lin Y-R., Muehlegger E., Shoag D., Gonza'lez M.C., Lazer D. (2015). Tracking employment shocks using mobile phone data. *Journal of the Royal Society Interface*, 12: 20150185.

Vicente M. R., López-Menéndez A. J., Pérez R. (2015). Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? Technological Forecasting and Social Change, Volume 92, March, Pages 132-139.

EK

Tablo 5 ARIMA Model Tahmin Sonuçları

15 Yaş Üstü İşsizlik Oranı					
	Google Index'in Olmadığı Model	"eleman arıyor" ifadeli Google Index	"iş arama" ifadeli Google Index	"iş bulma" ifadeli Google Index	"iş ilanı" ifadeli Google Index
C	2.312226*** (0.061691)	2.310403*** (0.064000)	2.287798*** (0.058979)	2.245919*** (0.055658)	2.223667*** (0.079955)
AR(1)	0.940734*** (0.031239)	0.940490*** (0.031323)	0.938568** (0.030995)	0.926268*** (0.032787)	0.941789*** (0.030655)
MA(1)	0.545754*** (0.075077)	0.547144*** (0.076059)	0.623694*** (0.070973)	0.701355*** (0.070983)	0.503228** (0.073092)
MA(2)	0.510996** (0.066570)	0.512992*** (0.067896)	0.585129** (0.071311)	0.655981*** (0.071224)	0.485176** (0.066569)
Eleman		0.000469 (0.004628)			
İsarama			0.007270* (0.003960)		
İsbulma				0.018396*** (0.004996)	
İsilanı					0.022798 (0.014515)
AIC	-4.802077	-4.787974	-4.822781	-4.874871	-4.805186
R <sup>2</sup>	0.973353	0.973356	0.973965	0.975677	0.973785
$\sum \varepsilon^2$	0.061279	0.061272	0.059870	0.055934	0.060285
15-24 Yaş Arası İşsizlik Oranı					
	Google Index'in Olmadığı Model	"eleman arıyor" ifadeli Google Index	"iş arama" ifadeli Google Index	"iş bulma" ifadeli Google Index	"iş ilanı" ifadeli Google Index
C	2.987147*** (0.046653)	3.005432*** (0.057303)	3.030985*** (0.050911)	3.030117*** (0.051334)	2.929373*** (0.082661)
AR(1)	0.491941*** (0.139335)	0.482173*** (0.137802)	0.427729*** (0.126238)	0.454000*** (0.143620)	0.481469*** (0.140536)
AR(2)	0.399019*** (0.140000)	0.409582*** (0.138510)	0.463695*** (0.118839)	0.436891*** (0.140547)	0.409760*** (0.142186)
MA(1)	0.873663*** (0.110466)	0.892852*** (0.107434)	0.966528*** (0.085713)	0.923004*** (0.107210)	0.877689*** (0.113933)
MA(2)	0.632295*** (0.063747)	0.632673*** (0.063016)	0.711202*** (0.054386)	0.670643*** (0.060781)	0.627118*** (0.065235)
Eleman		-0.004848 (0.005544)			
İsarama			-0.013105*** (0.004410)		
İsbulma				-0.012180** (0.005645)	
İsilanı					0.014942 (0.015883)
AIC	-4.506679	-4.498920	-4.550534	-4.515624	-4.499158
R <sup>2</sup>	0.952342	0.952652	0.955182	0.953502	0.952650
$\sum \varepsilon^2$	0.081374	0.080844	0.076525	0.079393	0.080848