



## Veri Madenciliği ve Turizmde Veri Madenciliği Çalışmaları\*\* (Data Mining and Data Mining Studies in Tourism)

\* Tuğba ŞEN KÜPELİ <sup>a</sup> , Kurban ÜNLÜÖNEN <sup>a</sup> 

<sup>a</sup> Ankara Hacı Bayram Veli University, Faculty of Tourism, Department of Tourism Management, Ankara/Turkey

### Makale Geçmişi

Gönderim Tarihi:25.01.2021

Kabul Tarihi:15.02.2021

### Anahtar Kelimeler

Bilgi ve iletişim teknolojileri

Büyük veri

Veri madenciliği

Yapay zekâ

Makine öğrenmesi

### Öz

Modern teknoloji, veri üretme, depolama ve analizi konularındaki anlayışı internetin de etkisiyle hızla dönüştürmektedir. Farklı mecralarda oldukça büyük miktardaki veriler olağanüstü bir hızda ve çeşitlilikte üretilmekte, aynı hızla veri depolama ve analiz tekniklerine olanak sağlayan teknolojik çözümler geliştirilmektedir. Bu çalışma kapsamında ise bilgi ve iletişim teknolojilerinin gelişimine bağlı olarak ortaya çıkan ve kullanımı hızla tüm alanlara yayılan veri madenciliği ve turizm alanındaki gelişimi incelenmektedir. Bu amaçla, veri madenciliği ve birbirleriyle ilişkili karmaşık yapıdaki diğer kavramlar açıklanıp, veri madenciliği teknikleri kullanılarak yapılan 1999 ile 2020 yılları arasındaki çalışmalar incelenmiştir. Buna göre, veri madenciliği konusundaki çalışmaların, turizm pazarlaması, imaj yönetimi veya turizm talebi gibi farklı turizm dinamiklerine yönelik çözümler getirmesine rağmen, yeterli düzeyde olmadığı ve kavramın alandaki gelişiminin eksik kaldığı sonuçlarına ulaşılmıştır. Çalışmada, veri madenciliği konusuyla ilgili genel bir çerçeve çizilerek turizm literatüründeki mevcut durumun aktarılması bakımından bu alanda önemli bir ihtiyaç giderilmeye çalışılmaktadır.

### Keywords

Information and communication technologies

Big data

Data mining

Artificial intelligence

Machine learning

### Abstract

With the power of the internet, digital technology is increasingly changing the perception of data creation, storage, and analysis. A large volume of data is generated at an incredible speed and variety in several different media, and at the same speed, technological solutions are created that enable data storage and advanced analysis techniques. The concept of data mining, which emerge due to the advancement of information and communication technologies and its development in the tourism field is examined. Thus, data mining and other related concepts are explained, and studies conducted using data mining techniques between 1999 and 2020 examined. Accordingly, while solutions were proposed for various tourism dynamics, including tourism marketing, demanding or image management in data mining studies, it was concluded that they were not at a high extent and that the progress of the conceptualization in the field remained insufficient. This research, therefore, is one of the leading studies that attempt to meet the needs in terms of drawing a general framework about the concept of data mining and including the current developments in tourism literature.

### Makalenin Türü

Araştırma Makalesi

\* Sorumlu Yazar

E-posta: tugba.sen@hbv.edu.tr (T. Şen Küpeli)

DOI: 10.21325/jotags.2021.789

\*\*Bu makale, Prof. Dr. Kurban Ünlüönen danışmanlığında yürütülmekte olan Tuğba Şen Küpeli'nin Doktora tez çalışmasından üretilmiştir.

## **GİRİŞ**

Bilgi ve iletişim teknolojilerinin ilerlemesi ile internetin yaygınlaşması; küresel çapta, büyük miktarlarda, yüksek hız ve çeşitlilikte dijital veriler üretilmesini, dolayısıyla üretilen verileri depolama, işleme, analiz etme ve yönetmeyi sağlayacak büyük veri uygulamalarının gelişmesini sağlamıştır (Akıncı, 2019, s. ii; Ersöz, 2019, s. 13). Modern bilgisayar sistemleri oldukça büyük miktarlardaki çeşitli türden verileri, bankamatiklerden uzaydaki uydulara kadar farklı birçok kaynaktan toplayabilmektedir (Bramer, 2007, s. 2). Öyle ki tahminlere göre 2025 yılına kadar, çeşitli platformlarda depolanacak toplam küresel veri miktarı 200 zettabayt (1 zettabayt = 1,000,000,000,000= 10<sup>12</sup> gigabayt olarak hesaplanmaktadır) seviyesine ulaşacaktır. Dahası 2030 yılına kadar tahminen 7,5 milyar insan çevrimiçi olacak ve veri üretecektir (Morgan, 2020, s. 1). Giderek büyüyen ve yaygın olarak kullanılabilen, devasa büyüklükteki bu veriler, “veri çağı” olarak adlandırılan yeni bir dönemin oluşmasına neden olmaktadır (Han, Pei & Kamber, 2011, s. 2).

Büyük miktardaki verilerin üretimi 1960’lı yıllara kadar uzanmasına rağmen, büyük verinin dikkat çekmeye başlaması, depolama ve analiz teknolojilerinin gelişmesiyle olmuştur. Astronomi ve genom gibi bilimlerin 2000’li yıllarda gelişmesi de “büyük veri” teriminin türetilmesini tetiklemiştir (Jeffcock, 2018; Mayer-Schönberger & Cukier, 2013, s. 14). Veri analizinin giderek önem kazanmasıyla, verinin içinden saklı olan bilgileri elde edebilmek kritik olarak görülmeye başlamıştır. Öyle ki bu bilgiler; bir işletmenin faaliyetlerini sürdürmesinde, farklı bilim dallarında önemli keşiflerin yapılmasında, çeşitli hastalıkların nedenlerinin ve tedavilerinin tanımlanmasında veya hava durumu ve doğal afetlerin tahmininde kullanılabilir (Bramer, 2007, s. 2). Böylece sektörler büyük veriyle yeniden şekillenerek veri değer oluşturmak, inovasyon yapmak, yeni ürün geliştirmek ve rekabet avantajı sağlamak üzere kullanılan, hayati öneme sahip bir hammaddeye dönüşmektedir (Mayer-Schönberger & Cukier, 2013, s. 11-13). Bu nedenle çeşitli kamu kurumları, işletmeler ve diğer kuruluşlar, amaçlarına bağlı olarak farklı türde verileri depolamaktadırlar (Dolgun, Özdemir & Oğuz, 2009, s. 48). Veri madenciliği teknikleriyle de verilerdeki izlerin peşine düşülerek, belirli sonuçlar çıkarılmaya ve özetlenmeye çalışılmaktadır (Mayer-Schönberger & Cukier, 2013, s. 16).

Nitekim veri tek başına anlamlı olmamakla birlikte, verinin işlenmesini olanaklı kılacak depolama ve analiz yöntemlerinin gelişmesiyle, veriden çeşitli çıkarımlar elde edilerek, iş süreçlerinde daha etkin kararlar verilebilecektir. Böylece, tüketici istek ve ihtiyaçlarını hızlı ve doğru bir biçimde karşılayacak mal ve hizmetler geliştirilebilecek, işletme maliyetleri azaltılacak ve rekabet avantajı sağlanacaktır. Bu nedenle, veri madenciliği konusunda yapılan çalışmalar eğitimden sağlık sektörüne, tarımdan finansa ve bilişimden ulaştırmaya kadar pek çok alanda uygulayıcıların yanında akademik alanda da giderek yaygınlaşmaktadır. Bu alanlardan birisi de turizmdir (Ersöz, 2019, s. 14-17; SAS, 2020).

Her ne kadar turizm alanında yapılan çalışmalar giderek artış gösterse de veri madenciliği konusu henüz yeni olarak nitelendirildiğinden, diğer alanlarda olduğu gibi yapılacak çalışmalara hala ihtiyaç duyulmaktadır. Veri madenciliği konusunu, teknik yönüyle ele alan çalışmaların literatürde ağırlık göstermesi (Bach, Schatten ve Marušić, 2013, s. 28; Li, Li, Zhang, Hu ve Hu, 2019, s. 20) ve sektör verilerinin daha çok “araç” olarak ele alınmasına bağlı olarak kavramın turizm literatüründeki teorik gelişimi de eksik kalmaktadır. Böylece, veri madenciliği kavramının turizm bağlamında kapsamlı bir şekilde ele alınması ihtiyacı ortaya çıkmaktadır.

Bu ihtiyaçtan hareketle mevcut alıřmada; veri madencilięi kavramı ile etkileřimde olduęu bilgi ve iletiřim teknolojileri, b y k veri ve yapay zek  gibi kavramlar aıklanarak, mevcut turizm literat rindeki alıřmalar incelenmiřtir. Bu sayede, veri madencilięi teknikleriyle turizm dinamiklerine y nelik getirilen  z mler hakkında genel bir perspektif sunulmaktadır. Dahası, karmařık yapıdaki birbirleriyle iliřkili dięer kavramlar aıklanarak da bu alanda arařtırma yapanlara temel nitelikte bir kaynak oluřturulmaya alıřılmıřtır.

## **Literat r**

Bu b l mde, veri madencilięinin anlařılabilmesi iin kavramın ortaya ıkmasına neden olan; bilgi ve iletiřim teknolojileri, b y k veri, veri depolama ve analizi ile yapay zek  gibi kavramlar birbirleriyle olan etkileřim ve turizm baęlamında aıklanacaktır.

## **Bilgi İletiřim Teknolojileri**

İletiřim teknolojilerin temeli basılı kitaplardan bařlamak  zere, sırasıyla gazete, telefon, radyo, televizyonun yaygınlařmasıyla ilerlemiř, 1970'li yıllara gelindięinde hızlı bir ivme kazanarak, uydu yayınları, video oyunları, cep telefonları ve sonrasında internet ile farklı bir boyut kazanmıřtır (Zhu, 2018, s. 11). Artan dijitalleřme, eřitli t rde ve  zellikte bilgi (veri) oluřturulan ve iletilen cihazlar (akıllı saatler, akıllı ev robotları, arabalar vb.) ile mecralar (youtube, facebook, instagram vb.) t reyerek bilgi ve iletiřim teknolojilerinin i ie gemesini ve geliřimini saęlamıřtır (Ghasemaghahi, Ebrahimi & Hassanein, 2018, s. 102-102; Grant & Meadows, 2018; Morgan, 2020, s. 2).

Oluřan muazzam hacim, hız ve eřitlięe sahip veriler, iřletmelerin karar alma performanslarının iyileřtirilmesini saęlama potansiyeline sahip olan, b y k veri devrimine sebep olarak g sterilmektedir. Dolayısıyla, b y k verinin anlamlı olması ve rekabet avantajı oluřturacak řekilde iřlenebilmesi iin veri analitięi teknik ve aralarının kullanımı artırılmakta, veri analitięi giderek  rg tsel bilgi teknolojisi haline gelmektedir (Ghasemaghahi & Calic, 2020, s. 147; Ghasemaghahi vd., 2018, s. 101-102). Bilgi teknolojilerinin kullanımıyla, t keticilerin istek ve ihtiyaları birebir pazarlama y netimi ile karřılanabilmekte, dahası m řteri tatmini ve baęlılıęı  st d zeye ıkarılabilmektedir (D. Buhalis, 1998, s. 419). M řteriler ise mal ve hizmetler hakkında daha fazla bilgi edinebilmekte, satın alma veya kullanma deneyimlerini de dięer insanlarla dijital teknolojiler sayesinde hızlı ve kolay bir řekilde paylařabilmektedir (Pertheban, Narayana Samy & Shanmugam, 2020, s. 2-3).

Dięer sekt rlerde olduęu gibi, bilgi iletiřim teknolojileri, turizm sekt r n  de k resel apta d n řt rmektedir. Artan talep, deęiřen turistik beklentiler ile turistlerin tatmin edilebilmesi iin iřletmeler, bilgi iletiřim teknolojilerine duyarsız kalamayarak, operasyonel ve stratejik y netim alanlarında bir d n ř m gerekleřtirmeye y nelik faaliyetlerde bulunmaktadırlar (Law, Buhalis & Cobanoęlu, 2014, s. 728).  rneęin, bir turist tatilini planlama ařamasında, dięer turistlerin paylařtıęı seyahat deneyimlerini g zden geirerek, seyahat planını edindięi bilgilerle řekillendirebilmektedir (Dimitrios Buhalis & O'Connor, 2005, s. 728; Pertheban vd., 2020, s. 2-3). Turistler bilgi ve iletiřim teknolojilerini; bilgi arayıřı, alternatiflerin deęerlendirilmesi, satın alma kararının verilmesi, satın alma sonrasında kullanabilmektedirler. Sosyal medya ise evrimii pazarlamada ve turistlerin karar vermelerinde  nemli bir rol oynamaktadır (Law vd., 2014, s. 729-732). Dahası, TripAdvisor veya Yelp gibi yorum siteleri; marka bilinci oluřturabilecek, iřletme itibarını ve iř performansını etkileyebilecek, evrimii kulaktan kulaęa (eWOM) iletiřimini kolaylařtırmaktadır (Dimitrios Buhalis, 2019, s. 268). Dolayısıyla, bilgi ve iletiřim teknolojilerinin, arz tarafında olduęu gibi, kaliteli hizmet elde etmek iin talep tarafında da kullanmasının kaınılmaz olduęu anlařılmaktadır.

Çevrimiçi mecralardaki etkileşimin giderek artması, milyonlarca insanın seyahatleri hakkındaki, analiz edilmesi ve anlamlı bütünler çıkarılması gereken büyük bir dijital veri gerçeğiyle karşılaşılmasına neden olmuştur (Mariani, 2019, s. 299; Mistilis & Dwyer, 2000, s. 58-59).

### **Büyük Veri (Big Data) Kavramı**

Büyük veri, her ne kadar 1990'lı yıllarda ortaya çıkmış olarak kabul edilse de büyük veri kümelerinin kökeni, veri merkezlerinin kurulduğu 1960'lara kadar uzanmakta, yaygınlığı ve önemi giderek artmaktadır (Firican, 2017; Oracle, 2020). Dahası, gelişen astronomi ve genom gibi bilimler 2000'li yıllarda “büyük veri” teriminin türetilmesini sağlamıştır (Mayer-Schönberger & Cukier, 2013, s. 14).

Kavramın ilk tanımı 2001 yılında, büyük verinin üç ana (3V) özelliği olan hacim (veri miktarı), hız (veri oluşturma, değiştirme ve aktarım hızı) ve çeşitliliği (verinin farklı formatlar/yapılarda olması) tanımlayan Doug Laney tarafından yapılmıştır (Mariani, 2019, s. 300). Ancak, büyük veri kavramının tanımlanmasına yönelik çeşitli görüşler yer almaktadır. Yapılan açıklamalar genel olarak, büyük verinin “ne” olduğu ve “ne işe yaradığı” üzerine odaklanan farklı iki bakış açısına göre şekillenebilmektedir (Gandomi & Haider, 2015, s. 138). Kavramın ne olduğuna odaklanan bir çalışmada büyük veri; tüm veri setine veya büyük bir bölümüne ulaşma durumu veya ana kitleyi temsil oranını ifade eden göreceli büyüklüktür (Mayer-Schönberger & Cukier, 2013, s. 14-36). Bir diğer tanımda, çeşitli kaynaklardan üretilen ve içinde bilgi barındıran çok sayıdaki farklı veri türüdür (Zheng, Zhu & Lyu, 2013, s. 403). Ne işe yaradığı konusuna odaklanan diğer bir tanımda büyük veri, gelişmiş iç görü ve karar verme için uygun maliyetli ve yenilikçi bilgi işleme biçimleri gerektiren büyük hacimli, yüksek hızlı ve çok çeşitli bilgi içeren veri bütünüdür (GartnerITGlossary, 2020). Büyük veri, geleneksel (ilişkisel ve/veya monolit) veri tabanı sistemleriyle depolanması, yönetimi ve analizi genellikle çok pahalı olan verilerdir (GoogleCloud, 2020). Benzer şekilde, büyük veri, boyutu veya türü geleneksel veri tabanı sistemlerinin yakalama, yönetme ve işleme yeteneğinin ötesinde olan veridir. Büyük veri tanımlanırken, taşıdığı özelliklere de odaklanılmaktadır. Büyük verinin temel özellikleri arasında 3V'si olarak ifade edilen büyük hacim (volume), yüksek hız (velocity) ve çok çeşitliliğe (variety) daha sonra değişkenlik (variability), gerçeklik (veracity), geçerlilik (validity), hassaslık (vulnerability), oynaklık (volatility), değer (value), doğrulama (verification) ve görselleştirme (visualization) özellikleri eklenerek (Firican, 2017) yapılan tanımlar da bu kavramlar etrafında şekillenmeye devam etmektedir.

Büyük veri kavramının tam tanımlanamaması, belki de sektör tarafından fazlaca ilgi görmesi ve önde gelen teknoloji firmalarının (Örneğin: IBM, Google, Tech America Foundation, Gartner IT Glossary) kendi bakış açılarına, sahip oldukları veri türüne veya özelliklerine göre büyük veriyi tanımlama çabalarından kaynaklanıyor olabilir. Öyle ki teknoloji firmaları genellikle faydacı bir yaklaşım göstererek “ne işe yarıyor” sorusuna cevap veren tanımlamalar yapmaya çalışırken, alandaki araştırmacılar daha çok “ne” sorusunu yanıtlayan tanımlar yapmaktadırlar. Öte yandan, dijital veri üretimi ve analizi konusunun hızlı değişkenlik özelliği, kavramın sürekli olarak yeniden değerlendirilmesini gerektirmektedir. Büyük veri içinde belki de keşfedilmesi gerekli olan birçok özellik barındırdığından, tanımını da zorlaştırmaktadır.

Büyük veriyi tanımlayan bu özelliklerin yanı sıra sağlıklı bir büyük veri değer ekosisteminde; yapılandırılmış, yapılandırılmamış, çok dilli, makine ve sensör tarafından oluşturulan, statik ve gerçek zamanlı olmak üzere farklı türde veriler yer almaktadır (Cavanillas, Curry & Wahlster, 2016, s. 8). Bu gibi özellikleri taşıyan verilerin üretildiği veri kaynaklarına örnek olarak; ortalama sayıları 2020 yılında 31 milyara ulaştığı tahmin edilen, arabalar, trafik

ışıkları, ev güvenlik sistemleri ve akıllı hoparlörler gibi cihazlar verilebilir (Morgan, 2020, s. 2). Dahası, bu gibi veri kaynakları, yapay zekâ (Artificial Intelligence [AI]), mobil cihazlar, sosyal medya ve nesnelerin interneti (Internet of Things [IoT]: birbirleriyle bağlantılı tüm alıcı ve cihazın oluşturduğu küresel ağ) tarafından yapılandırıldığından, geleneksel verilere göre daha karmaşık, gerçek zamanlı ve büyük ölçekte veri oluşturmaktadır (GoogleCloud, 2020; IBM, 2020). Dijitalleşme öylesine hızlı gerçekleşmektedir ki 2000’li yıllarda dünyada saklanan bilginin yalnızca %25’i dijital veriyken 2013 yılı için bu oran %90 üzeri olarak tahmin edilmektedir (Mayer-Schönberger ve Cukier, 2013, s. 17). Dahası, büyük veri üretimi hakkında yapılan tahminler, bilimi ve insanları sürekli olarak yanıltarak, hesaplamaların ötesine geçmektedir (Morgan, 2020). Uluslararası veri merkezinin yaptığı hesaplama göre 2010 yılında bir önceki yıla göre 0.8 zettabayt artış göstererek 1,2 (1,2 trilyon gigabayt) zettabayt dijital veri üretilmiştir. Bu artış oranından hareketle 2020 yılında üretilen dijital veri 35 zettabayt olarak hesaplanmasına rağmen bu rakama neredeyse 2018 yılında ulaşılmıştır (Press, 2020). Dahası 2019 yılında 45 zettabayt olan dijital veri miktarı, 2020 yılına gelindiğinde Covid-19 sonrası evden çalışma ve eğitim sebebiyle yaşanan dijitalleşme oranındaki artışın etkisiyle 59 zettabayt olarak tahminlerin çok üzerine çıkmıştır. Farklı kaynaklarda 2025 yılına kadar üretilen küresel veri miktarının ise 175 veya 200 zettabayt olacağı yönündedir (Holst, 2020; Morgan, 2020; Press, 2020; Reinsel, Gantz ve Rydning, 2018).

Bu öngörüler göstermektedir ki Covid 19’un neden olduğu etkiyle de işletmeler değişen rekabet şartlarına ayak uydurmak ve iş modellerinde, hizmet sunumlarında ve teknoloji kullanımlarındaki dönüşümü hızla gerçekleştirmek zorunda kalmaktadırlar. Araştırmalar, salgın sonrası bulut sistemlerine olan talebin 2020’de %13 arttığını, Mart 2020’de %17 olan yatırımların iki hafta içinde %30 yükseldiğini ve 2020 yılının ilk çeyreğinde Amazon web hizmetlerinin %33, Google bulut sistemlerinin gelirlerinin %34 oranında arttığını göstermektedir. Öte yandan, yapılan araştırmalara göre küresel çapta üretilen veri büyüklüğüne rağmen, karar verme ve iş süreçlerinde kullanımı oransal olarak daha azdır. Nitekim işletmelerde toplanılan veri miktarının neredeyse %68’i analize tabi tutulamamakta ve verinin içinde gömülü olan değerli bilgilerden bir fayda elde edilememektedir (SEAGATE, 2020, s. 14; Woodie, 2020a).

Büyük veri hayati bir ekonomik girdi ve yeni iş modellerinin temeli olarak, önemli bir dönüşüm gerçekleştirmektedir. 20.yy’da toprak ve fabrikalar gibi fiziksel varlıklardan, markalar ve fikri mülkiyet haklarına doğru kayan ‘değer’ artık “veri” ye kaymaktadır (Mayer-Schönberger & Cukier, 2013, s. 23). Büyük veri üretimine yönelik gelişmelerin, işletmeler açısından nasıl kullanışlı hale gelebileceği sorusuna verilebilecek temel cevap belki de tüketicilerin davranışlarına, muhtemelen haberleri bile olmadan, onların gözünden bakabilmek (Wilkinson, 2018, s. 166) olacaktır. Öyle ki cep telefonlarının, sosyal medya kanallarının ve kredi kartlarının her kullanıldığında veya çevrimiçi bir form oluşturulduğunda ve bir video izlendiğinde tüketici davranışını anlamaya yönelik olarak hükümetler veya işletmeler tarafından kullanılacak olan büyük miktardaki veri, kendiliğinden üretilmektedir (Blossom & Alex, 2018, s. 221-222). Dahası büyük veri tek başına bir anlam ifade etmezken, hızı ve çeşitliliği gibi özellikler işletme performansını etkilemede önemli faktörler olarak öne çıkmaktadır (Ghasemaghahi & Calic, 2020, s. 158).

Mevcut gelişmelere ilaveten, dünyanın yeni doğal kaynağı olarak görülen verinin; depolanması, işlenmesi ve siber suçlara karşı korunmasına yönelik çalışmaların yürütülmesine de ihtiyaç duyulmaktadır (Morgan, 2020, s. 2).

## Veri Depolama ve Analizi Teknolojisi

Teknolojik gelişmeler, birçok alanda olduğu gibi, veri üretme, depolama ve analizi konularında da hızlı bir dönüşüm ve gelişime yol açmaktadır. Öyle ki modern bilgisayar teknolojilerindeki ilerlemeyle birlikte çok büyük miktarlarda çeşitli türde veriler dijital ortamlarda giderek daha düşük maliyetlerle depolanabilmektedir (Bramer, 2007; Han vd., 2011; Mayer-Schönberger & Cukier, 2013). Dahası veri üretme hızı, yapılan tahminleri aşarak hayal edilmesi güç boyutlara ulaşabilmektedir. Örneğin, 2025 yılına kadar Dünyadaki veri miktarının 175 zetabayt olacağı öngörüsünden hareketle üretilecek tahmini verinin büyüklüğü algılanamayacak boyut kazanmaktadır (Reinsel vd., 2018, s. 3,6). Bu nedenle büyük veri kavramının gelişiminde, depolama teknolojileri oldukça önemli bir yere sahiptir. Nitekim veri üretiminin anlamlı olabilmesinin ilk adımı depolamadır. Xu (2014, s. 33)'ya göre veri depolama teknolojisinin gelişmesiyle 1990 ve 2013 yılları arasında depolama maliyetleri yıllık %33 oranında azalmıştır ve her geçen gün bu teknoloji ilerlemektedir. Pazarın bir kısmı sabit disk sürücülerindedir ve sırasıyla Seagate (%41), Western Digital (38) ve Toshiba (%21) hâkim firma konumundadır. Bu şirketler gelecek 10 yıl içinde 50TB (Terabayt), sonrasında 100TB bellek kapasiteli sürücüler geliştirilebileceğini savunmaktadırlar (Woodie, 2020b). Sabit disk sürücülerine alternatif olarak, kullanıcılarına depolamaya ilave farklı çözümler sunabilen bulut tabanlı sistemler geliştirilmiştir (Mazumdar, Seybold, Kritikos & Verginadis, 2019). Bu yeni dijital depolama teknolojileri, (örneğin: NoSQL), verinin saklanması yanı sıra ölçeklenebilir, güvenilebilir, yönetilebilir, verimli ve esnek veri modelleri geliştirilebilmesini sağlamaktadır. Google Dosya Sistemi, (GFS), Hadoop Dağıtım Dosya Sistemi (HDFS), BigTable, HBase ve Scalaris gibi farklı tasarım amaçlarına ve faydaya yönelik birçok depolama teknolojisi geliştirilmiştir (Siddiqa, Karim & Gani, 2017). Bulut tabanlı sistemler kullanıcılarına, verilerin depolanması, işlenmesi ve analizinin, uygun maliyet ve esnek iş modelleriyle, daha ölçeklenebilir şekilde yapılmasını vadederek işletmelerin ilgisini çekmeye çalışmaktadırlar (GoogleCloud, 2020).

Böylece büyük veri kümelerini işleyebilen bu teknolojik ilerleme ve maliyet etkinliğinden tüm sektörler etkilenebilmektedir. Örneğin, sağlık sektöründe Çin hükümeti büyük veri yönetimini etkin bir şekilde kullanarak, Covid-19 salgınının yayılmasını engellemek ve kontrol altına alınmasını sağlamak için kullanmıştır (Wu, Wang, Nicholas, Maitland & Fan, 2020, s. 1). Medya sektöründe sosyal medyanın analizi, çok sayıda bireyin oluşturduğu haberleri özetleyerek, gazeteciliği dönüştürme potansiyeline sahiptir. Taşımacılık sektöründe, ulaşım sistemlerinin veri yönetimi; yolcuların deneyimini artırabilirken aynı zamanda karar vericilerin şehir trafiğini daha iyi yönetmelerine yardımcı olabilmektedir (Strohbach, Daubert, Ravkin & Lischka, 2016, s. 122). Benzer gelişmeler turizm sektöründe de yaşanmaktadır. Tüketiciler kendi deneyimlerini farklı birçok mecra (sosyal medya, bloglar, rezervasyon siteleri veya yorum siteleri vb.) üzerinden paylaşmakta, tatil planlarını da paylaşılan bu deneyimlere göre şekillendirmektedirler (Dimitrios Buhalis, 2019; Law vd., 2014; Mariani, 2019, s. 299; Mistilis & Dwyer, 2000, s. 58-59). Öte yandan verinin işlenebilmesiyle, veri ancak karar alma süreçlerinde etkin bir rol alan araca dönüşebilmektedir. Büyük veri analitiği bu noktada devreye girerek, yapay zekâ gibi dijital teknolojilerin gelişmesini de tetiklemekte, tüketicilerin geçmiş ve güncel davranışlarının öğrenilmesi ve turistlerin ihtiyaçlarının belirlenmesine olanak tanımaktadır (Mariani, 2019, s. 301).

## Yapay Zekâ (Artificial Intelligence-AI)

Karar verme, insanların sıklıkla yaptıkları ve etkili sonuçları olan ve bazen karmaşık bir aktivite olduğundan, araştırmacılar, karar verme yeteneklerini artırmak üzere bilgisayar teknolojileri geliştirmişlerdir. Yapay zekâ

alanındaki gelişmeler, çeşitli uygulamalarla bu amacın gerçekleştirilmesini sağlamaktadır (Phillips-Wren, 2012, s. 1). Yapay zekâ, akıllı donanım veya yazılım üretimi dâhilinde, öğrenme ve problem çözme gibi insan davranışlarını kopyalayabilen bilgisayar bilimi alanıdır. John McCarthy tarafından 1956'da kullanılmasıyla beraber akademik bir disiplin olarak kabul edilerek, giderek çeşitli şekillerde insan gibi davranan robotları veya yazılımları tasvir eden yaygın bir kavram haline gelmiştir (Liu, 2020). İlk yapay zekâ çalışmalarından biri, Gary Kasparov ile IBM'in yazılımı olan Deep Blue satranç maçı olarak kabul edilmektedir. Dünyanın en büyük satranç oyuncularından biri olan Kasparov, 1997 yılında yapay zekâyâ karşı yenilmesiyle, yapay zekâ insanların rekabet edemeyeceği bir teknoloji olarak görülmeye başlanmıştır (Oracle, 2018).

Yapay zekâ ile bütünleşmiş karar verme destek sistemleri (AI-integrated decision making support) veya akıllı karar destek sistemleri (intelligent decision support systems [IDSS]); finans, sağlık hizmetleri, pazarlama, ticaret, siber güvenlik ile komuta ve kontrol gibi alanlarda karar almaya yardımcı olmak için giderek daha fazla kullanılmaktadır (Phillips-Wren, 2012, s. 1). Öyle ki bir hisse senedinin performansına ilişkin tahminler Twitter'ın çözümlenmesiyle, bazı ürün tavsiyeleri Amazon, Netflix kullanıcıların site içi etkileşimlerinin çıkarılmasıyla ve bireysel tercihler Facebook ve LinkedIn üzerinden yapılan sosyal ağ analiziyle ortaya çıkarılmaktadır (Mayer-Schönberger & Cukier, 2013, s. 22). Dahası "Siri", "Spotify" gibi sık kullanılan uygulamalarla günlük hayatın bir parçası olarak gelişimine devam etmektedir (Oracle, 2018). Bu gelişimle, 2019'da teknoloji şirketlerinin yapay zekâ yazılımlarına yaptıkları harcamaların 13,5 milyar ABD doları olduğu tahmin edilmektedir (Liu, 2020).

### **Veri Madenciliği (Data Mining) Kavramı**

Veri miktarlarındaki artışla beraber, verilerden anlamlı bilgilerin elde edilmesini sağlayacak analiz tekniklerini içeren veri madenciliği kavramı ortaya çıkmıştır (Mitchell, 1999, s. 1). Kavram literatürde, veri tabanlarından bilgi keşfi (knowledge discovery from databases) gelişmiş veri analizi (advanced data analysis), bilgi madenciliği (knowledge mining) ve makine öğrenmesi (machine learning) gibi adlarla tanımlanmaktadır (Altunkaynak, 2019, s. 13; Mitchell, 1999, s. 1). Gelişimi 1980'li yıllara dayanan makine öğrenmesi, yapay zekânın bir alt kümesi olarak insan beyninin çalışma prensibini taklit ederek, verilerden anlam çıkarmayı sağlayan teknikler bütünüdür (Jeffcock, 2018).

Bilinen istatistiksel ve ekonometrik yöntemlerle veri analizinde amaç; tahmin, özetleme, değerlendirme/keşif ve hipotez testi olmakla birlikte veri madenciliğinde tahmin, veri özetleme ve veriden farklı örüntüler çıkarma temel amaçlardandır ve analiz dahilinde doğru model bilinmemekle birlikte araştırmanın amacı doğru modeli keşfetmektir (Shapoval, Wang, Hara & Shioya, 2018; Varian, 2014).

Veri madenciliği teknolojisi web aramalarından, sosyal ağlarda içerik filtrelemeye, e-ticaret web sitelerindeki önerilerden, kameralar ve akıllı telefonlar gibi ürünlerde giderek daha fazla yer almaktadır (LeCun, Bengio & Hinton, 2015, s. 436). Örneğin, herhangi bir sosyal medya aracını kullanma esnasında, beğeni yapma, fotoğraf paylaşma, arkadaşlık ağlarını genişletme veya yorum yapma sonucunda "dijital iz" bırakılmaktadır. Veri madenciliği teknikleriyle bu izlerin peşine düşülerek, belli sonuçlar çıkarılmaya ve özetlenmeye çalışılmaktadır (Mayer-Schönberger & Cukier, 2013, s. 16). Veri madenciliği kullanılarak veriler; sınıflandırma, sayısal tahmin, ilişkilendirme ve kümeleme olmak üzere dört temel şekilde incelenmektedir (Bramer, 2007, s. 4). Farklı mecralarda oluşan veri çeşidine bağlı olarak, veri madenciliğiyle ilişkili çeşitli alt alanlar ortaya çıkmıştır. Yapılandırılmış veya yapılandırılmamış veriler için uygulanan; metin analizi (metin madenciliği), ses analizi, video analizi, gen analizi,

web analizi, sosyal medya analizi ve tahmine dayalı analiz türleri ve bu analizleri yapabilen farklı araçlar geliştirilmektedir (Altunkaynak, 2019, s. 15; Gandomi & Haider, 2015). Verilerin işlenmesi ve analiz edilmesi süreçlerinde özel yazılım programları kullanılmaktadır (Ersöz, 2019, s. 13). Veri madenciliği ile türetilen ilişkiler ve özetlere genellikle model veya kalıplar denir. Lineer denklemler, kümeler, grafikler, ağaç yapıları ve zaman serilerindeki tekrarlayan modeller veri madenciliği analiz teknikleri kullanılarak elde edilebilmektedir (Hand, Mannila & Smyth, 2001, s. 6). Yapılan analizler verinin elde edildiği mecra ve veri yapısına göre değişebilmektedir. Örneğin; lojistik regresyon kullanılarak yapılan basit bir makine öğrenimi algoritmasıyla bir gebenin sezaryenle doğuma uygunluğu değerlendirilebilir veya sinir ağları Naive Bayes adı verilen makine öğrenimi algoritması, yasal e-postayı, spam e-postadan ayırabilir (Bengio, Goodfellow & Courville, 2017, s. 2). Turizm alanında da farklı amaçlarla çeşitli sonuçlara ve çözümlere ulaşılabilmektedir. Turistlerin seyahat davranışlarıyla bir destinasyona ilişkin mevsimsellik, turistlerin konaklama, yeme içme veya ziyaret edilen cazibe unsurlarına yönelik tercihleri makine öğrenmesi teknikleriyle öğrenilerek, pazarlama stratejilerinde gerekli adımlar atılmasını kolaylaştırabilmektedir. Dahası turist profillerinin ayrıntılı incelenmesi sayesinde, etkili pazar bölümlendirme ve turist istek ve ihtiyaçlarına yönelik doğru ürün tasarımı yapılabilmesine, müşteri tatmini ve tekrar ziyaret isteği oluşturacak kalite standartları oluşturulabilmesine olanak sağlanabilmektedir (Emel & Taşkın, 2006; Olmeda ve Sheldon, 2002; Tichter, Grossmann & Werthner, 1999).

Veriden anlamlı ve faydalı bilgi çıkarma süreci, belirli bir maliyet, zaman ve emek gerektirmektedir. Nitekim büyük veri söz konusu olduğunda, verinin tamamından faydalı bilgiler elde etmek her ne kadar istenen sonuç olsa da mevcut uygulamalarda, toplam verinin ancak %20-30'u kullanılabilir (SEAGATE, 2020, s. 14; Woodie, 2020a). Geriye kalan veri, işletmeler için bir fırsat veya harcanacak çeşitli maliyetler göz önünde bulundurulduğunda riskli bir yatırım olarak da görülebilir. Bu durum her işletmenin veri kaynağı ve yapısıyla beraber değişebilmektedir.

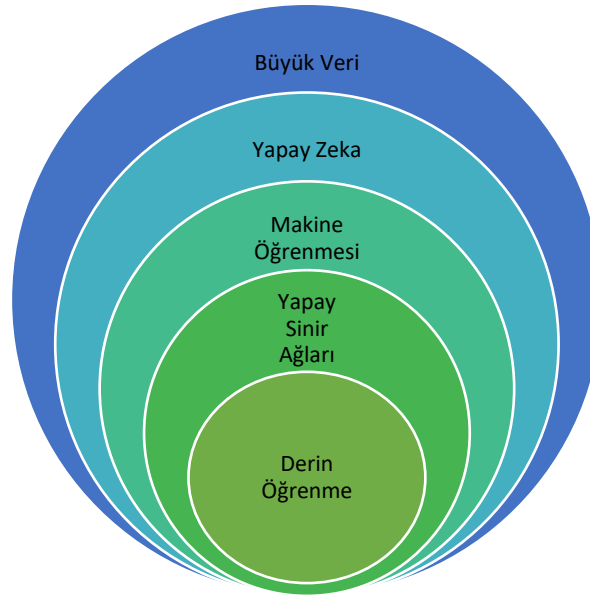
Makine öğrenmesinde kullanılan yapay nöronlar (sinir ağları) karmaşık problemlere çözüm üretebilme performansı göstermekle beraber, nesnelere tanıma gibi basit beyin fonksiyonlarını gerçekleştirme ve doğal verileri işleme yetenekleri açısından sınırlı kalmıştır. Analiz anlayışının gelişmesiyle beraber derin öğrenme kavramı ortaya çıkmıştır. Derin öğrenmede insan beyninin öğrenme prensibi taklit edilmekte ve makine öğrenmesine göre daha fazla sinir ağı yapısı kurulmaktadır (Jeffcock, 2018).

### **Derin öğrenme (Deep Learning)**

Yapay sinir ağları olarak adlandırılan ve insan beyninin işleyişini taklit eden algoritmaları içeren, yapay zekâ ile makine öğreniminin bir alt bölümü olarak 2010 yılından sonra gelişme göstermiştir. Derin öğrenme, bilgisayarların deneyimden öğrenmesini ve kelimeleri belli bir hiyerarşik düzende anlamasını sağlayan, mantıksal yapılar kurabilen makine öğrenmesi algoritmasıdır. Otonom arabalardan, konuşma, ses veya görsel tanıma gibi, gerçekleştirilmesi daha önce mümkün olarak görülmeyen, makine öğrenmesine göre daha karmaşık problemlerin çözümlenmesinde kullanılmaktadır (Goodfellow, Bengio, Courville & Bengio, 2016, s. 351; Jeffcock, 2018). Derin öğrenmenin, yapay sinir ağı adı verilen ve insan beyninin biyolojik sinir ağından esinlenerek oluşturulan katmanlı algoritma yapısı, bir insanın çevresindeki verileri kullanarak çıkarımlarda bulunmasına benzer bir mantıkla, denetimli veya denetimsiz öğrenme yoluyla çalışmaktadır. Bu sayede standart makine öğrenimi modellerinden çok daha yetenekli bir öğrenme süreci gerçekleştirebilmektedir (Wolfewicz, 2020). Görüntülerdeki nesnelere tanımlamak, konuşmayı metne dönüştürmek, haber öğelerini, gönderileri veya ürünleri kullanıcıların ilgi alanlarıyla eşleştirmek ve ilgili arama



sonuçlarını seçmek gibi giderek kullanımı artan çeşitli uygulamalar, derin öğrenme tekniğiyle yapılandırılmaktadır (Acumen, 2020; LeCun vd., 2015, s. 436). Örneğin derin öğrenmeyle, bir fotoğrafta insan mı, bir araba mı, yoksa başka bir nesne mi olduğu ile tahmin edilen şeyin görünüşüyle ilgili detaylar belirlenebilmektedir (Bengio vd., 2017, s. 6)



**Şekil 1.** Kavramlar Arası İlişki

**Kaynak:** Bu şekil oluşturulurken Kavlakoglu (2020) ve Wolfewicz (2020)'in aktardıkları şekil ve bilgilerden yararlanılmıştır.

Mevcut çalışmada açıklanmaya çalışılan kavramların daha iyi anlaşılabilmesi için Şekil 1 incelendiğinde, her biri kendisinden öncekinin bileşeni olan, birbiriyle bağlantılı kavramlardan bahsedildiği görülebilir. Özetle, büyük veri, yapay zekâ uygulamalarının gelişmesini tetikleyerek, önce makine öğrenmesi ve sonra derin öğrenmeye uzanan bir bilgisayar teknolojisinin temelini oluşturmaktadır. Nitekim teknolojik öğrenme süreci algoritmaları besleyen veri büyüklüğüne bağlı olarak görülmektedir (Wolfewicz, 2020).

Derin öğrenme teknolojisindeki gelişmeler, turizm sektörüne yönelik de çeşitli fırsatlar barındırmaktadır. Tekrarlayan sinir ağı (recurrent neural network-rnn), uzun-kısa süreli bellek (long-short-term memory -lstm), dikkat mekanizması (attention mechanism), derin sinir ağları (deep residual neural network- ResNet) ve evrimsel sinir ağları (convolutional neural networks- CNN/ConvNet) gibi modeller, daha doğru turizm talep tahmininin yapılmasında (Law, Li, Fong ve Han, 2019, s. 411), turistlerin seyahatlerine ilişkin yorumlarının sınıflandırılmasında (Martín, Torres, Aguilar ve Diaz, 2018) veya bir destinasyonun imaj analizinde kullanılabilir (Sheng, Zhang, Shi, Qiu & Yao, 2020). Mevcut potansiyeline bağlı olarak derin öğrenme pazarı her geçen gün gelişmektedir. Nitekim küresel derin öğrenme pazarının, 2026 yılına kadar 56 milyon doları bulacağı tahmin edilmektedir (Acumen, 2020).

### Araştırmanın Yöntemi

İşletmeler, turistlerin istek ve ihtiyaçlarını büyük veri teknolojileri kullanarak daha iyi anlayabilmektedir (Dimitrios Buhalis & O'Connor, 2005, s. 12). Büyük verinin, veri madenciliği teknikleriyle işlenebilmesi için gelişen teknolojiler, işletmelere ve araştırmacılara farklı fırsatlar sunmaktadır. Nitekim veri madenciliği henüz gelişmekte olan bir alan olmakla birlikte, işletmeler için de bir sermaye kaynağı olarak görülmektedir.

Veri madenciliği kavramına olan sektör ilgisiyle beraber akademik çalışmalar da ilgi artmaktadır. Böylece mevcut çalışmada, veri madenciliği konusunda turizm alanında yapılan çalışmaların taraması yapılarak, akademik yayınların yıllara göre dağılımı ve hangi turizm dinamiklerine yönelik olduğu gibi sorulara cevap aranmıştır. Böylece, mevcut araştırmada turizm alanında yapılan çalışmalar incelenerek konunun alandaki gelişimi ortaya konulmaya çalışılmıştır. Bu amaçla Scopus veya Web of Science gibi spesifik veri tabanları yerine, daha kapsamlı tarama yapıldığı kabul edilen (Alaei, Becken ve Stantic, 2019, s. 181) “Google Akademik” üzerinden 06.01.2021 tarihinde bir tarama yapılmıştır. Tarama sırasında üç aşamalı bir sınırlama uygulanmıştır. Buna göre ilk aşamada, “data mining” (veri madenciliği) kavramı aratılmıştır. İkinci aşamada, “data mining tourism” kelimeleri birlikte aratılmıştır. Son olarak üçüncü aşamada, başlık sınırlaması yapılarak “data mining tourism” kelimeleri aratılmıştır. Bu taramayla birlikte aşağıdaki sorulara yanıt verilmeye çalışılmıştır:

- Veri madenciliği konusu hakkındaki çalışmaların sayısı kaçtır?
- Veri madenciliği konusunda turizm alanında yazılan makale sayısı kaçtır?
- Başlığında “veri madenciliği ve turizm” kelimeleri yer alan, kaç çalışma vardır?
- Çalışmalarda kullanılan anahtar kelimeler nelerdir?
- Turizm alanındaki veri madenciliği çalışmalarının uygulandığı yerler nerelerdir?
- Turizm alanındaki veri madenciliği çalışmalarında kullanılan veri miktarı nedir?
- Turizm alanındaki veri madenciliği çalışmaları hangi turizm dinamiklerine yöneliktir?
- Turizm alanındaki veri madenciliği çalışmalarında kullanılan analiz teknikleri nelerdir?

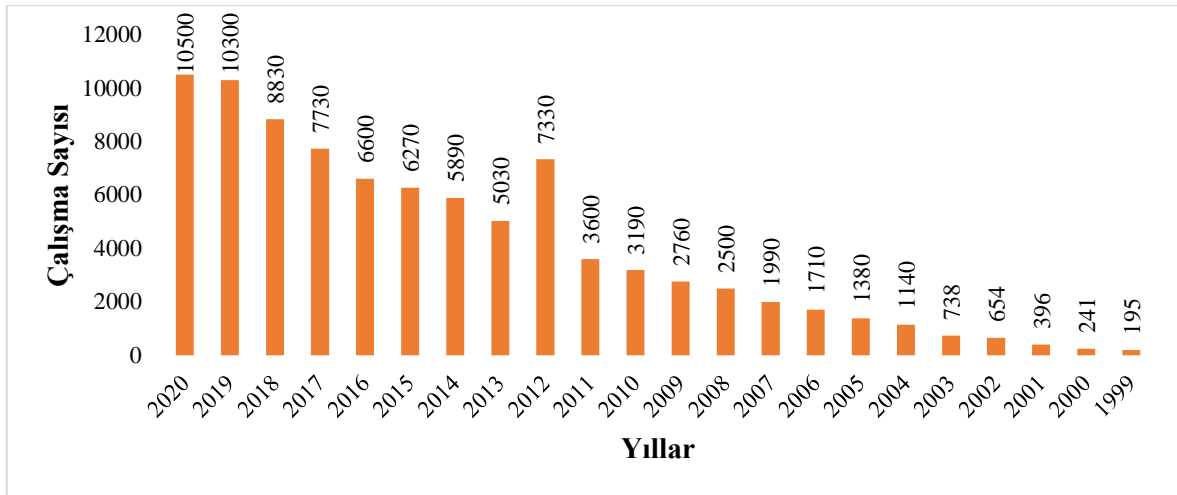
Makalelere yönelik yukarıdaki sorulara yanıtlar bulgular kısmında verilmektedir.

## Bulgular

Yapılan taramalardan ilk aşamada sonucuna, data mining (veri madenciliği) kavramı hakkında, 2 milyon 250 bin çalışmaya ulaşılmıştır. İkinci aşamada, “data mining tourism” kelimeleri birlikte aratıldığında, 79 bin 200, (Türkçe sayfalarda 229) çalışmaya ulaşılmıştır.<sup>1</sup> Yapılan bu taramayla, veri madenciliği ve turizm kelimelerinin birlikte yer aldığı çalışma sayısının, genel çalışma sayısına oranının yalnızca %3,5 olduğu söylenebilir. Bu da turizm alanında daha fazla çalışma yapılması gerektiğini göstermektedir. Buradan hareketle, Grafik 1’de “data mining tourism” kelimeleri çalışmaların herhangi bir yerinde olacak şekilde aratıldığında, yıllara göre yapılan çalışma sayıları gösterilmektedir.

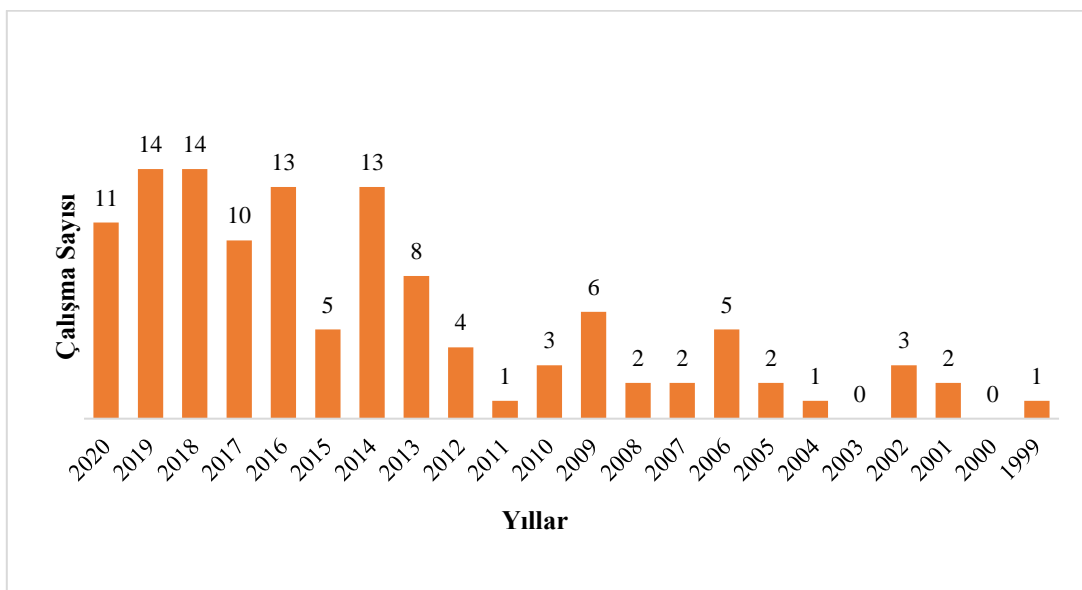
Öte yandan taramada ortaya çıkan rakamlar ve makaleler, veri madenciliği konusunda yapılan tüm çalışmaları değil yalnızca bir yansımasını göstermektedir. Nitekim veri tabanlarında aratılan kelimelerden herhangi birinin çalışmada yer alması durumunda elde edilen rakama yansımakla birlikte, bazen bu çalışmalar ilgili konu hakkında olmayabilmektedir. Bu durumun tespit edilmesinden dolayı üçüncü aşamada yeni bir sınırlama yapılarak tarama gerçekleştirilmiştir. Tarama, “makale başlığında” olacak şekilde (Google Akademik üzerinden tarih, dergi, dil, yazar ve başlığa göre tarama yapılabilmektedir) sınırlandırıldığında 126 (Türkçe sayfalarda 1) çalışmaya ulaşılmaktadır. Yıllara göre tek tek arama yapıldığında toplam sayı 120 olmuştur.

<sup>1</sup> Bu taramalar genel eğilime örnek göstermesi bakımından yapılmış olup, diğer veri tabanlarında arama yapıldığında farklı rakamlara ulaşılması muhtemeldir.

**Grafik 1.** Turizmde Veri Madenciliği Çalışmalarının Yıllara Göre Dağılımları

**Kaynak:** Google Akademik (2020).

İkinci aşamadaki taramaya göre, veri madenciliği konusunda yapılan ilk çalışmalar 1999 yılındadır. Başlık sınırlaması getirilerek “hotel data mining” kelimeleri birlikte aratıldığında ilk çalışmaya 1998 (Ha ve Park, 1998) yılında, “hospitality data mining” kelimeleri aratıldığında ise 1998 (Law, 1998) yılında rastlanmıştır. Bu çalışmalar dışında, daha önceki bir tarihte herhangi bir çalışmaya rastlanmamıştır. Böylece turizm alanına yönelik veri madenciliğine olan ilginin 90’lı yılların sonunda başladığı ve yıllara göre artış yaşandığı söylenebilir. Üçüncü aşamada yapılan taramadaki (data mining tourism) veri madenciliği çalışmalarının yıllara göre sayısı Grafik 2’de verilmiştir. Grafikte, 2021 yılına ait veri, mevcut çalışmadaki taramanın Ocak 2021’de yapılması nedeniyle gösterilmemiştir. Sonuçlara göre, 2000 ve 2003 yıllarında çalışmaya rastlanmazken, 2011 yılında yalnızca bir ve 2015 yılında ise beş çalışmaya rastlanmıştır.

**Grafik 2.** Başlığında Turizm ve Veri Madenciliği Olan Çalışmaların Yıllara Göre Dağılımları

**Kaynak:** Google Akademik (2020).

Tarama sonucunda elde edilen 1999 yılından 2020 yılına kadar yayımlanmış 120 çalışmaya ait anahtar kelimelerin kelime bulutu hazırlanmıştır. Şekil 2’de görüleceği üzere, çalışmalarda en fazla veri, turizm, madencilik,

analiz, web, büyük, ticaret, yönetim, bilgi, sektör, model, algoritma, metin, yöntem, davranış, destinasyon, Müşteri ilişkileri yönetimi (Crm), duyu, pazarlama kavramları yer almaktadır.



**Şekil 2:** Turizm Alanındaki Çalışmalarda Kullanılan Anahtar Kelimeler

Elde edilen tüm çalışmalara yönelik (120 çalışma) diğer detaylara; basılı veya elektronik kopyasının yer almaması, erişim sağlanamaması, birden fazla defa sayılması veya taramada yer almasına rağmen turizm alanına yönelik olmaması gibi nedenlerle yer verilememiştir. Böylece söz konusu makalelerden, turizm alanıyla doğrudan ilgili olup, ulaşılabilen çalışmalardan bazıları Tablo 1’de özetlenerek genel çerçeve çizilmeye çalışılmıştır. Nitekim bu çalışmanın amacı alanda yapılan tüm çalışmaları özetlemek veya sınıflandırmaktan daha çok genel bir perspektif sunarak, bu alanda araştırma yapanlara temel nitelikte bir kaynak oluşturmaktır.

**Tablo 1.** Turizm Alanında Veri Madenciliği Konusunda Yapılan Çalışmalar

Çalışma	Uygulama yeri	Veri Miktarı	İlgi alanı	Kullanılan/ Bahsedilen Analiz teknikleri
<i>Tichter vd. (1999)<sup>1</sup></i>	Tux, Avusturya	**	Turist davranışları	Kümeleme, log-lineer modeller ve karar ağaçları
<i>Olmeda ve Sheldon (2002)</i>	*	*	Turizm pazarlaması, müşteri ilişkileri yönetimi	*
<i>Emel ve Taşkın (2006)</i>	Bursa, Türkiye	204	Pazar bölümlendirme	Sınıflandırma
<i>Law, Mok ve Goh (2007)</i>	*	*	Turizmde talep tahmini	***
<i>Salguero, Araque, Carrasco, De Campos ve Martínez (2007)</i>	Granada çevresi	ve Hava durumu tahmin verileri	Rekreatif bir spor olan yelken uçuşu için doğru şartların belirlenmesi	Fuzzy SQL
<i>Carrasco, Araque, Salguero ve Vila (2008)</i>	**	**	Turizm aktivitesine yönelik uygun coğrafi alanı ve müşteri özelliğini belirleme	SQL Yapısal Sorgulama Dili (Structured Query Language - SQL). Bulanık Yapısal Sorgulama Dili (FSQL).

**Tablo 2.** Turizm Alanında Veri Madenciliği Konusunda Yapılan Çalışmalar (Devamı)

<i>Okuhara, Yeh, Hsia ve Ishii (2008)</i>	<b>Turistik alan</b>	<b>*</b>	<b>Turizm için coğrafi erişilebilirlik</b>	<b>Mixture of Expert (MoE)</b>
<i>Guoxia ve Jianqing (2009)</i>	Guilin, Çin	50 farklı alandan turistlerle ilgili bilgileri (yaş, cinsiyet, kalış süresi vb) içeren 5500 kayıt	Guilin şehrine gelen turistlerin profilleriyle, deneyim arayışları sınıflandırılması	SAS/EM
<i>(Koh ve Gan, 2009)</i>	Singapur	300 turistten elde edilen anket verileri	Sağlık Turizmde hastaların, hastane seçim ve deneyimlerinin kategorilendirilmesi	Kümeleme analizi
<i>Mirela, Tiberiu ve Adina (2009)</i>	Bucovina, Romanya	10 otele ait müşteri veri tabanı	Stratejik pazarlama	-
<i>Bach vd. (2013)</i>	Web of science, SCOPUS	Web of Science and Scopus 88 makaledeki anahtar kelimeler	1995 ve 2013 yılları arasında Turizm dergilerinde yayımlanan çalışmalarda veri madenciliğini konu alan çalışmaların anahtar kelimelerin belirlenmesi	Kavramsal ağ analizi Wordle1 ve LaNet-vi2 araçları kullanılarak yapılmıştır.
<i>Gu, Zhang, Chen ve Chang (2016)</i>	Shenzhen, Çin	Çin'in Shenzhen şehrindeki 86 turistik cazibe merkezinden toplanan yaklaşık 340.000 check-in kaydı	Turist davranışları, faaliyetleri ile turist tercihlerini belirlemek ve destinasyon olarak Shenzhen şehrinin problemlerinin tespiti	Sina micro-blog location-based service (LBS)
<i>Shapoval vd. (2018)</i>	Japonya	Japonya Seyahat Bürosu'ndan alınan 4000 anket verileri	Turistlerin tercihlerini ve deneyimleri sonucu tekrar ziyaret niyetini belirlemek	Karar Ağaçları (Decision Trees-DTs)
<i>Li vd. (2019)</i>	Google Akademik Science Direct ACM Dijital Kitaplığı Citeseer Kütüphanesi Web of Science IEEE explore Springer Link	İlgili veri tabanlarında, 2014-2019 yılları arası Turizm alanında metin madenciliği üzerine yayımlanmış çalışmalar	Turizm metin veri madenciliği teknikleri ve uygulamalarının sistematik bir incelemesi	Doğal dil işleme (natural language processing-NLP)
<i>Halkiopoulou ve Koumparelis (2020)</i>	Yunanistan	1883-2019 dönemini kapsayan 117 otel ilanı ve 1907-2019 dönemini kapsayan 59 acente reklamı	Otel ve Acentaların reklamlarında odaklandıkları noktaları belirlemek ve zaman içindeki değişimi incelemek	Google Vision API
<i>Park, Xu, Jiang, Chen ve Huang (2020)</i>	Jeonju, Gangneung ve Chuncheon, Güney Kore	Turistlerin cep telefonlarından elde edilen 116,807 konuma ilişkin veriler	Destinasyonlardaki yeni "sıcak nokta/ popüler alan (hot spots) ve bu yerlerdeki mekânsal etkileşimleri ortaya çıkarmak	density-based spatial clustering of applications with noise (DBSCAN)
<i>Yan, Chen ve Wang (2020)</i>	Lombok ve Bali, Endonezya	Lombok adası için 2,007,802 Tweet Bali adası için 7,643,592 Tweet kaydı	Ağustos 2018 Endonezya depremi sonrasında Twitter kullanıcıların bu destinasyona yönelik yorumlarının analizi	Duygu analizi (sentiment analysis) ve Gizli Dirichlet Ayırımı (Latent Dirichlet allocation-LDA)
<i>Zhang ve Dong (2021)</i>	Çin	Bir navigasyon uygulaması işlevi gören Beidou uygulamasından kişisel bilgiler ve turist rehberlerinin tur programlarına ait bilgiler ve 100 turiste uygulanan anket	Görüntü izleme sistemi ile kullanıcılar için kişiselleştirilmiş seyahat rotası oluşturabilme.	Field Programmable Gate Array -FPGA

**Not:** <sup>1</sup> : Makalenin bir kısmına ulaşılabilmektedir. \*: Makalede uygulama yapılmamıştır. \*\* Bilgiye ulaşılamamıştır. \*\*\*NN (SFP) = Denetimli/Öğretimli İleri Beslemeli Sinir Ağı; NN (BP) = Geri yayımlı Sinir Ağı; RS = Kaba Kümeler; NN (Elman's) =

Elman'ın Yapay sinir ağı; NN (MLP) = İleri Beslemeli Çok Katmanlı Algılama Sinir Ağı; NN (RBF) = Radyal Temelli Fonksiyon Sinir Ağı; NN (BAYN) = Bayes Sinir Ağı; SPR = Destek Vektör Regresyonu; GMDH = Grup Metoduyla Bilgi İşleme; FTS = Bulanık Zaman Serileri; GFM = Gri Tahmin Modeli.

Tichter vd., (1999)'in yaptığı çalışma alanda ilk kabul edilebilecekler arasındadır. Yerel turizm eğilimlerinin analiz edilmesi için Avusturya'nın Tux beldesine ait turizm verileri elde edilmiştir. Veri madenciliği kullanılan çalışmada, kümeleme, log-lineer modeller ve karar ağaçları gibi çok değişkenli keşifsel istatistiksel teknikler uygulanarak, turistlerin davranış kalıplarındaki mevsimsellik araştırılmıştır.

Diğer bir çalışmada (Olmeda ve Sheldon, 2002) veri madenciliği tekniklerinin turizmin internetten pazarlanmasında ve elektronik müşteri ilişkileri yönetiminde potansiyel kullanımları değerlendirilmiştir. Müşteri profili oluşturma, sorgulama, yönlendirme, e-posta filtreleme, çevrimiçi müzayedeler ve e-katalogların güncellenmesi makalede açıklanmaktadır. Ayrıca, web sayfalarından bilgi getirmesi, bilgilerin heterojen özellikte olması, gizlilik prensipleri, video kalitesini etkileyen kısıtlı bant genişliği, depolama ve mobil cihazların görüntüleme ve güç kullanımında sınırlı olması gibi veri madenciliği tekniklerinin uygulanmasındaki zorluklara değinilmiştir.

Emel ve Taşkın (2006)'ın çalışmasında, bir iç turizm pazarının bölümlerinin tahminlenmesi için veri madenciliği tekniklerinden C5 sınıflandırma algoritması kullanılmıştır. Bursa'daki seyahat acentalarından anket yöntemiyle toplanan 204 soru formu elde edilmiştir. Müşteri profiline yönelik 17 soruya ilaveten, konaklama türü ve seçim davranışına yönelik 32 sorudan alınan veriler analiz edilmiştir. Yerli turistlerin davranış kalıplarını tanımlamak üzere konaklama süresi, aylık gelir ve yaş gibi özelliklere göre pazar bölümleri ve bu turistlerin aradıkları tatil ve otel özellikleri çıkarılmıştır.

Law vd., (2007), mevcut çalışmaya benzer bir şekilde, turizm talep tahmininde veri madenciliğine ilişkin 1980 ile 2007 yılları arasında olmak üzere EBSCO veri tabanından, ScienceDirect ve Hospitality & Tourism Complete indeksinde taranan dergilerde, turizm talep tahmini üzerine yapılmış toplam 174 makaleyi incelemişlerdir. Yayımlanan ilk makale 1999 yılında olmak üzere 14 makalede veri madenciliği tekniklerinin kullanıldığı tespit edilmiştir.

Salguero vd., (2007) çalışmalarında hava sıcaklık tahminlerine yönelik farklı kaynaklardan alınan bilgilerle, motorsuz hava uçaklarıyla süzülme için en uygun hava koşullarına karar verecek bir algoritma üzerine çalışmışlardır. Çalışmada, veri madenciliği teknikleriyle seyahat acentaları veya tur operatörlerinin müşterilerine önercekleri rekreatif faaliyetler hakkında karar vermeleri için ihtiyaç duydukları bilgilerin keşfedilebileceği sonucuna ulaşılmıştır.

Carrasco vd., (2008)'nin kitap olarak yayımladığı çalışmaya sınırlı bir erişim sağlanmıştır. Buna göre planör olarak bilinen motorsuz uçakların uçurulduğu rekreatif bir aktivite için doğru coğrafi konuma karar verme sürecin karmaşık olduğu belirtilmiştir. Çalışmada, konum seçim sürecinde, çevresel faktörler ve müşteri özelliklerinin göz önüne alınabildiği bir model önerisi için veri madenciliği teknikleri (FSQL) uygulaması yapılmaktadır.

Guoxia ve Jianqing (2009)'in çalışması, Guilin şehrinde daha fazla konferans aktivitesi yapılması gerektiğini ve alışveriş mekânlarının zayıf olmasının, zengin cazibe merkezinde bile kötü bir algıya neden olabileceğini göstermektedir. Buna ilaveten, Turizm ürünlerinin, sadece cazibe merkezlerine odaklanması, turizm ürününün kendi değerini düşürebilir.

Koh ve Gan (2009)'nın çalışmalarında sağlık turizmi için gelen turistlere yapılan uygulama sonucunda, turistler farklı kümelere ayrılmışlardır. Buna göre, uzmanlık, teknoloji ve kolaylık gibi özellikler arayanlar özel hastaneleri tercih ederken, yüksek beklentisi olanlar kamu hastaneleri tercih etmektedirler. Özel hastanelerdeki yabancı hastalar, kamu hastanelerindekilere kıyasla, hastane personelinin tutumu açısından daha büyük bir beklenti deneyimine (ve dolayısıyla daha düşük memnuniyet) sahiptir.

Mirela vd., (2009) çalışmalarında farklı veri tabanlarından elde edilen karmaşık bilgilerin tek bir çatı altında toplanarak işlenebildiği veri madenciliği modeli oluşturmayı amaçlamışlardır. Çalışmada, dağınık verilerin veri madenciliği teknikleriyle analizinin mümkün olduğu ve bu tekniklerin stratejik pazarlama amaçlarını gerçekleştirebilmek üzere, turizm alanında kullanılabileceği sonucuna ulaşılmıştır.

Bach vd., (2013)'nin çalışmasında Scopus ve Web of Science veri tabanlarında yer alan turizmle ilgili bilgi ve iletişim teknolojileriyle (BİT) ilgili dergilerdeki makalelerin anahtar kelimeleri analiz edilmiştir. Dergi araması yapılırken, veri madenciliği, bilgi keşfi, turizm, turist, destinasyon, seyahat ve otel kelimeleri kullanılmıştır. Bu sınıflandırmaya uyan 1995 ve 2013 yılları arasındaki 88 makale incelenmiştir. Sonuçlara göre, BİT dergilerinde öne çıkan kelimeler sistem, turizm, veri, tavsiye eden, veriler ve bulanık kelimeleriyken; turizmle ilgili dergilerde önemli anahtar kelimeler, bölümlendirme, analiz, madencilik, turizm ve veridir. Bu sonuç, BİT dergilerinde yazılımların, turizm dergilerinde ise verinin kendisinin üzerinde daha fazla durulduğunu göstermektedir.

Gu vd., (2016) çalışmalarında turistlerin giriş, dolaşım ve kalış süreleri gibi verileri analiz ederek, Çin'in Shenzhen şehrinin bir destinasyon olarak nasıl görüldüğü ve problemleri tespit edilmeye çalışılmıştır. Tema parkları yapılarak 1990'lı yıllarda cazibe merkezi olarak pazarlanmaya çalışılan şehirde, turistlerin bazen, trafik yoğunluğu ve uzun bekleme sıraları nedeniyle çekim merkezine gelmelerine rağmen giriş yapamadıkları tespit edilmiştir. Buna ilaveten gidilen cazibe merkezlerdeki tercih önceliği, kalış süreleri ve daha çok tercih edilen merkezlerin hangileri olduğu da çalışmada analiz edilenler arasında yer almaktadır.

Shapoval vd., (2018) bir varış noktası olarak Japonya'ya gelen turistlerin önemli deneyimlerini belirlemek; turistlerin neleri beğenilerini ve beğenmediklerini; bu deneyimlerin ve tercihlerin tatmini nasıl etkilediğini ve bir varış noktası olarak Japonya'ya tekrar ziyaret niyetini belirlemek üzere yapılmıştır.

Li vd., (2019) kapsamlı bir literatür taraması gerçekleştirmişlerdir. Çalışmada turizm alanında yapılarak yayımlanmış olan makaleler çeşitli veri tabanlarından elde edilerek incelenmiş ve hangi büyük veri analiz yöntemini kullandıkları, uygulama alanları, verilerin nerelerden elde edildiği ve alana katkısı gibi özelliklere göre sınıflandırılmıştır. Çalışma sonuçlarına göre mevcut literatür, daha çok veri ile algoritma üzerine odaklanmakta, turistlerin davranışları ile sektörün yapısı göz önünde bulundurulmadan veri analizi ve hizmete yönelik değerlendirmeler yapılmaktadır. Bu nedenle, turizm alanına özgü değişkenlerin dikkate alındığı çalışmaların yapılmasına ihtiyaç duyulduğu aktarılmaktadır.

Halkiopoulos ve Koumparelis (2020)'un Yunanistan'daki otel ve seyahat acentalarının tanıtımlarının temel amaçlarını, zaman içindeki değişimlerini ve metinsel ve görsel unsurlarını incelemek üzere bir çalışma yapmışlardır. Çalışma kapsamında, 1883-2019 yılları arasında 117 otel ilanı ve 1907-2019 yılları arasındaki 59 acente reklamı incelenmiştir. Nesneleri ve yüzleri algılayan, basılı ve el yazısı metinlerini okuyabilen ve değerli meta veriler oluşturan ilgili makine öğrenimi algoritmaları sağlayıcısı olarak Google Vision API kullanılarak analiz yapılmıştır.

Sonuç olarak, otel ve acenta reklamları beş farklı aşamada tarihsel bir evrim geçirmiştir. Geline son aşamada, dijital dönüşümle, mobil ve internet üzerinden reklam sayıları artarken dergi ve gazete reklamları yıllar içinde azalmıştır.

Park vd., (2020) Güney Kore'deki üç farklı destinasyonu ziyaret eden yabancı turistlerin bir yıl boyunca GSM verileri işlenmiştir. Bu analizlerle, her bir destinasyona yönelik olarak; turistlerin hareketli olduğu saat aralıkları, buralardaki kalış süreleri, gittikleri yerler (semt pazarları, üniversiteler ve kültürel miras unsurları) ve ortalama günlük yürüme mesafeleri gibi bilgiler elde edilmiştir. Bu bilgiler, her bir destinasyondaki turistlerin keşfedilmemiş eğilimlerini ortaya çıkarmada önemli bir adım oluşturmakta, pazarlamacılar için gerçek zamanlı bilgiler edinilmesini sağlamaktadır. Sonuç olarak, bir destinasyonun rekabet gücünü artırabilecek sürdürülebilir ve yenilikçi destinasyon planlaması ve yönetimi tasarlamak için bu gibi bilgilerin önemine bu çalışmada değinilmektedir.

Yan vd., (2020) çalışmalarında Ağustos 2018 depremi sonrası, Endonezya turizmine yönelik Twitter kullanıcılarının yaptıkları yorumlardaki duyguları, zaman serisi ile birlikte analiz etmeyi amaçlamışlardır. Analiz sonuçları depremden hemen sonra olumsuz yorumlar artmış olsa da sonraki aylarda bu sayıda azalma olmuştur. Diğer araştırmalardan farklı olarak, fiziksel iyileştirme çabaları karşısında insanların duygularındaki iyileşmelerin seyri ortaya çıkarılmaya çalışıldığı savunulmaktadır. Çalışma deprem sonrası Balı ve Lombok adalarının imaj ve itibar yönetimi ile turizm tanıtımında kullanılabilir faydalı bilgiler sunduğu ifade edilmektedir.

Zhang ve Dong (2021), çalışmalarında bir navigasyon sistemi ile farklı kaynaklardan alınan verileri işleyerek kişiye özel tur programları ve anlık öneri sunan bir uygulama yapmaya çalışmışlardır. Çalışma sonucunda, turistlerin ziyaret ettikleri destinasyon ve tercihleri ile kişisel özellikleri, bir sonraki ziyaret edecekleri destinasyondaki tercihlerini tahmin etmede kullanılabilir. Yazarlar, farklı kanallardan üretilen büyük verilerin analizi sayesinde, turist tercihleri tahmin edilebilir ve bir turizm destinasyonu için marka imajı tasarımı, marka imajı iletişimi, marka imajı koruma, marka imajı değerlendirme gibi uygulamaların başarılı olarak yürütülebilmektedir.

## **Sonuç**

Bir destinasyona geri dönme niyeti, ziyaretçilerin mal ve hizmetlerinden duydukları tatmin düzeyine bağlı olarak görülmektedir (Shapoval vd., 2018, s. 23). Bu istenen sonuca ulaşmak için merkezi ve yerel yönetimler ve özeld turistik işletmeler (oteller, restoranlar, havayolları, seyahat acentaları vb.) çeşitli stratejiler geliştirmeye çalışmaktadırlar. Bunlar arasında veri madenciliği, kullanım alanı artarak gelişme gösteren bir kavram olarak ortaya çıkmaktadır. Nitekim dijitalleşmenin neden olduğu oldukça büyük miktarlardaki veri doğru analiz edildiğinde, sürdürülebilir ve yenilikçi turizm planlaması ve pazarlaması açısından çeşitli fırsatlar sunabilecek ve nihayetinde destinasyon rekabet gücünü artıracak potansiyelleri barındırmaktadır (Park vd., 2020, s. 14-15). Turizm sektörü de dijitalleşmenin etkisiyle çeşitli platformlarda üretilen farklı türde büyük verileri bünyesinde barındırmaktadır. Öyle ki sektörde ilki üreticiden tüketiciye, diğeri tüketiciden üreticiye olmak üzere iki farklı veri akış yönü vardır. Veri madenciliği teknikleri kullanılarak analiz edilebilen bu veriler, turist davranışları (Tichter vd., 1999), müşteri ilişkileri yönetimi (Olmeda ve Sheldon, 2002), talep tahmini (Law vd., 2007), turist tercihlerinin tahmini (Zhang ve Dong, 2021), destinasyonlara yönelik problemlerin belirlenmesi (Gu vd., 2016), Pazar bölümlerinin belirlenmesi, (Emel ve Taşkın, 2006), turistlerin tekrar ziyaret niyeti (Shapoval vd., 2018), destinasyon imaj yönetimi (Yan vd., 2020), cazibe merkezlerinin ve kalış sürelerinin belirlenmesi (Gu vd., 2016) ile turist profilinin analizi ve turist harcamalarının tahmini gibi konularda, klasik istatistiksel analizlerden daha kapsamlı analiz fırsatı sunabilmektedir (Bose, 2009, s. 936).



Bu çalışmada, “Google Akademik” sayfasında 1999 ile 2020 yılları arasında veri madenciliği konusunda yapılan çalışmaların dağılımları incelenerek, doğrudan turizm alanına yönelik olan bazı çalışmaların hangi turizm dinamiklerine çözüm getirdiği görülmüştür. Buna ilaveten turizm alanına yönelik yapılan veri madenciliği çalışmaları genel olarak iki kategoride değerlendirilebilir. Çalışmaların bir kısmını, turizm sektöründe üretilen büyük veri setlerini (turist veya geceleme sayıları, harcama miktarları, restoran yorumları, otel puanları vb.) temel alarak yapılan ve asıl amacın geliştirilen algoritmanın (yazılımların) veya analiz tekniğinin test edilmesi ve başarı oranlarının hesaplanmasıyla ilgili çalışmalar oluşturmaktadır. Diğer kategoride ise asıl amacın turizm sektörüne ve literatürüne katkı sağlaması olan ve kullanılan veri setinin yanında, yapılan analiz, yorum ve değerlendirmelerin de turizm alanına yönelik olan çalışmalar yer almaktadır. İkinci kategorideki çalışmaların sayısı kısmen azdır. Bu durum, Bach vd. (2013) ile Li vd., (2019)’nin çalışma sonuçlarında da gösterilmektedir. Bu kategoride kullanılan veri madenciliği tekniklerinin, turizm alanının literatürüne olduğu gibi turizm sektörünün gelişimine katkı sağlanması için hangi alanlarda, nasıl ve ne şekilde kullanılabileceği mevcut çalışmalarla örneklendirilmektedir.

Sonuç olarak, veri madenciliği kavramı ve yakın ilişkide olduğu; bilgi ve iletişim teknolojileri, büyük veri, yapay zekâ ve derin öğrenme gibi farklı diğer birçok kavramla akademik alanda olduğu gibi günlük hayatta da daha sık karşılaşılacaktır. Turizmde veri madenciliğine yönelik yapılan çalışmaların, örneklendirilen turizm dinamiklerine yönelik çeşitli çözümler sunmalarına rağmen, kavramın gelişimi için yeterli olmadığı söylenebilir. Özellikle turizm bakış açısı ve alan bilgisiyle birlikte analiz ve değerlendirme yapılan, veri madenciliği konusundaki çalışmalara ihtiyaç duyulmaktadır (Li vd., 2019). Mevcut araştırmada, sınırlı sayıda çalışma genel bir bakış açısıyla sunulmuş olup, daha kapsamlı çalışmalar yapılmalıdır.

## Beyan

Makalenin tüm yazarlarının makale sürecine verdikleri katkı eşittir. Yazarların bildirmesi gereken herhangi bir çıkar çatışması yoktur.

## KAYNAKÇA

- Acumen. (2020). Deep learning market global industry analysis, market size, opportunities and forecast, 2019 - 2026. *Acumen Research and Consulting*. Erişim Adresi: <https://www.acumenresearchandconsulting.com/deep-learning-market>. Erişim Tarihi: 05.01.2021.
- Akıncı, A. N. (2019). *Büyük veri uygulamalarında kişisel veri mahremiyeti* (Uzmanlık Tezi). TC Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı.
- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191.
- Altunkaynak, B. (2019). *Veri madenciliği yöntemleri ve r uygulamaları kavramlar – modeller–algoritmalar* (İkinci Baskı ed.). Ankara: Seçkin Yayınevi.
- Bach, M. P., Schatten, M., & Marušić, Z. (2013). Data mining applications in tourism: A keyword analysis. *Paper presented at the Central European Conference on Information and Intelligent Systems*. 24, 2013.

- Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning*. (1), USA: MIT press Massachusetts.
- Blossom, H. D., & Alex, N. (2018). Ehealth. In A. E. Grant & J. H. Meadows (Eds.), *Communication technology update and fundamentals* (16 ed.). New York: Routledge.
- Bose, I. (2009). Data mining in tourism. In *Encyclopedia of information science and technology, second edition* (pp. 936-940), IGI Global.
- Bramer, M. (2007). *Principles of data mining*. (180), Springer.
- Buhalis, D. (1998). Strategic use of information technologies in the tourism industry. *Tourism management*, 19(5), 409-421. doi:Doi 10.1016/S0261-5177(98)00038-7
- Buhalis, D. (2019). Technology in tourism-from information communication technologies to etourism and smart tourism towards ambient intelligence tourism: A perspective article. *Tourism Review*.
- Buhalis, D., & O'Connor, P. (2005). Information communication technology revolutionizing tourism. *Tourism recreation research*, 30(3), 7-16.
- Carrasco, R. A., Araque, F., Salguero, A., & Vila, M. A. (2008). Applying fuzzy data mining to tourism area. In *Handbook of research on fuzzy information processing in databases* (pp. 563-584): IGI Global.
- Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). The big data value opportunity. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New horizons for a data-driven economy* (pp. 3-11): Springer, Cham.
- Dolgun, M. Ö., Özdemir, T. G., & Oğuz, D. (2009). Veri madenciliğinde yapısal olmayan verinin analizi: Metin ve web madenciliği. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 2(2), 48-58.
- Emel, G. G., & Taşkın, Ç. (2006). Identifying segments of a domestic tourism market by means of data mining. In H.-D. Haasis, H. Kopfer, & J. Schönberger (Eds.), *Operations research proceedings 2005* (pp. 653-658), Springer.
- Ersöz, F. (2019). *Veri madenciliği teknikleri ve uygulamaları* (3. Baskı ed.). Ankara: Seçkin Yayınevi.
- Firican, G. (2017). *The 10 vs of big data*. Erişim Adresi: <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx>. Erişim Tarihi: 27.12.2020.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144. doi:10.1016/j.ijinfomgt.2014.10.007
- GartnerITGlossary (2020). *Big data*. Erişim Adresi: <http://www.gartner.com/it-glossary/big-data/>. Erişim Tarihi: 27.12.2020.
- Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of business research*, 108, 147-162. doi:10.1016/j.jbusres.2019.09.062
- Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *Journal of Strategic Information Systems*, 27(1), 101-113. doi:10.1016/j.jsis.2017.10.001
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*, (1), MIT press Cambridge.

- Google Akademik (2020). *Allintitle: Tourism "data mining"*. Erişim Adresi: [https://scholar.google.com.tr/scholar?q=allintitle:+tourism+%22data+mining%22&hl=tr&as\\_sdt=0,5](https://scholar.google.com.tr/scholar?q=allintitle:+tourism+%22data+mining%22&hl=tr&as_sdt=0,5). Erişim Tarihi: 06.01.2021
- GoogleCloud (2020). *Büyük veri nedir*. Erişim Adresi: <https://cloud.google.com/what-is-big-data?hl=tr>. Erişim Tarihi: 14.12.2020.
- Grant, A. E., & Meadows, J. H. (2018). *Communication technology update and fundamentals* (16 ed.). New York: Routledge.
- Gu, Z., Zhang, Y., Chen, Y., & Chang, X. (2016). Analysis of attraction features of tourism destinations in a mega-city based on check-in data mining—a case study of shenzhen, china. *ISPRS International Journal of Geo-Information*, 5(11), 210.
- Guoxia, Z., & Jianqing, T. (2009). The application of data mining in tourism information. *Paper presented at the 2009 International Conference on Environmental Science and Information Application Technology*.
- Ha, S. H., & Park, S. C. (1998). Application of data mining tools to hotel data mart on the intranet for database marketing. *Expert Systems with Applications*, 15(1), 1-31. doi:Doi 10.1016/S0957-4174(98)00008-6
- Halkiopoulou, C., & Koumparelis, A. (2020). Historical advertisements of hotels, tour agencies and involved services of tourism industry in greece: A data mining analysis via image processing. In V. Katsoni & T. Spyriadis (Eds.), *Cultural and tourism innovation in the digital era* (pp. 209-218): Springer.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*: Elsevier.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining (adaptive computation and machine learning)*, MIT Press.
- Holst, A. (2020). *Amount of information globally 2010-2024*. Statista. Erişim Adresi: <https://www.statista.com/statistics/871513/worldwide-data-created/#:~:text=The%20total%20amount%20of%20data,ever%2Dgrowing%20global%20data%20sphere>. Erişim Tarihi: 27.12.2020.
- IBM. (2020). *What is big data analytics?* IBM. Erişim Adresi: [https://www.ibm.com/analytics/hadoop/big-data-analytics?mhsrc=ibmsearch\\_a&mhq=Zettabyte](https://www.ibm.com/analytics/hadoop/big-data-analytics?mhsrc=ibmsearch_a&mhq=Zettabyte). Erişim Tarihi: 27.12.2020.
- Jeffcock, P. (2018). *What's the difference between ai, machine learning, and deep learning?* Erişim Adresi: <https://blogs.oracle.com/bigdata/difference-ai-machine-learningdeplearning#:~:text=Machine%20learning%20is%20a%20subset,to%20solve%20more%20complex%20problems>. Erişim Tarihi: 13.01.2021.
- Kavlakoğlu, E. (2020). *Ai vs. Machine learning vs. Deep learning vs. Neural networks: What's the difference?* Erişim Adresi: <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>. Erişim Tarihi: 27.05.2020.
- Koh, H. C., & Gan, L. L. (2009). A data mining analysis of healthcare tourism in singapore. *Advances on Information Processing and Management*, 26.

- Law, R. (1998). Hospitality data mining myths. *Hospitality Review*, 16(1), 7.
- Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 26(5), 727-750. doi:10.1108/Ijchm-08-2013-0367
- Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of tourism research*, 75, 410-423. doi:10.1016/j.annals.2019.01.014
- Law, R., Mok, H., & Goh, C. (2007). Data mining in tourism demand analysis: A retrospective analysis. *Paper presented at the International Conference on Advanced Data Mining and Applications*.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. doi:10.1038/nature14539
- Li, Q., Li, S., Zhang, S., Hu, J., & Hu, J. (2019). A review of text corpus-based tourism big data mining. *Applied Sciences*, 9(16), 3300. doi:ARTN 330010.3390/app9163300
- Liu, S. (2020). *Artificial intelligence (ai) worldwide - statistics & facts*. Erişim Adresi: <https://www.statista.com/topics/3104/artificial-intelligence-ai-worldwide/>. Erişim Tarihi: 27.12.2020.
- Mariani, M. (2019). Big data and analytics in tourism and hospitality: A perspective article. *Tourism Review*, 75(1), 299-303.
- Martín, C. A., Torres, J. M., Aguilar, R. M., & Diaz, S. (2018). Using deep learning to predict sentiments: Case study in tourism. *Complexity*, 2018, 1-9.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data, a revaluation that will transform how we live, work, and think* (B. Erol, Trans.), Houghton Mifflin Harcourt Publishing Company.
- Mazumdar, S., Seybold, D., Kritikos, K., & Verginadis, Y. (2019). A survey on data storage and placement methodologies for cloud-big data ecosystem. *Journal of Big Data*, 6(1), 15. doi:ARTN 1510.1186/s40537-019-0178-3
- Mirela, D., Tiberiu, S., & Adina, B. (2009). Toward a distributed data mining system for tourism industry. *Annals of Faculty of Economics*, 4(1), 922-926.
- Mistilis, N., & Dwyer, L. (2000). *Information technology and service standards in mice tourism*. Paper presented at the Journal of Convention & Exhibition Management.
- Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, 42(11), 30-36. doi:Doi 10.1145/319382.319388
- Morgan, S. (2020). *The 2020 data attack surface report arcsolve*. Erişim Adresi: <https://1c7fab3im83f5gqiow2qqs2k-wpengine.netdna-ssl.com/wp-content/uploads/2020/12/ArcserveDataReport2020.pdf>. Erişim Tarihi: 26.12.2020.
- Okuhara, K., Yeh, K.-Y., Hsia, H.-C., & Ishii, H. (2008). Data mining of geographical accessibility for tourism in regional area. *Paper presented at the 2008 Eighth International Conference on Intelligent Systems Design and Applications*.

- Olmeda, I., & Sheldon, P. J. (2002). Data mining techniques and applications for tourism internet marketing. *Journal of Travel & Tourism Marketing*, 11(2-3), 1-20.
- Oracle. (2018). *Make ai collaboration your next b2b marketing move*. Erişim Adresi: <https://blogs.oracle.com/cx/make-your-next-b2b-marketing-move-ai-collaboration>. Erişim Tarihi: 12.01.2021.
- Oracle. (2020). *What is big data? Big data defined*. Erişim Adresi: <https://www.oracle.com/big-data/what-is-big-data.html>. Erişim Tarihi: 03.01.2020.
- Park, S., Xu, Y., Jiang, L., Chen, Z., & Huang, S. (2020). Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data. *Annals of Tourism Research*, 84, 102973. doi:ARTN 10297310.1016/j.annals.2020.102973
- Pertheban, S., Narayana Samy, G., & Shanmugam, B. (2020). A systematic literature review: Information accuracy practices in tourism. *Journal of Quality Assurance in Hospitality & Tourism*, 21(1), 1-30. doi:10.1080/1528008x.2018.1563016
- Phillips-Wren, G. (2012). Ai tools in decision making support systems: A review. *International Journal on Artificial Intelligence Tools*, 21(02), 1240005. doi:Artn 124000510.1142/S0218213012400052
- Press, G. (2020). 6 predictions about data in 2020 and the coming decade. *Forbes*. Retrieved from <https://www.forbes.com/sites/gilpress/2020/01/06/6-predictions-about-data-in-2020-and-the-coming-decade/?sh=5d7138624fc3>
- Reinsel, D., Gantz, J., & Rydning, J. (2018). The digitization of the world: From edge to core. International Data Corporation. Retrieved from: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>. Erişim Tarihi: 21.12.2020.
- Salguero, A., Araque, F., Carrasco, R., De Campos, A., & Martínez, L. (2007). *Using fuzzy data mining for finding preferences in adventure tourism*. *Eurofuse Workshop*. Erişim Adresi: [https://www.researchgate.net/profile/Luis\\_Martinez19/publication/241881953\\_Using\\_Fuzzy\\_Data\\_Mining\\_for\\_finding\\_preferences\\_in\\_adventure\\_tourism/links/582446ef08aeb45b58899315/Using-Fuzzy-Data-Mining-for-finding-preferences-in-adventure-tourism.pdf](https://www.researchgate.net/profile/Luis_Martinez19/publication/241881953_Using_Fuzzy_Data_Mining_for_finding_preferences_in_adventure_tourism/links/582446ef08aeb45b58899315/Using-Fuzzy-Data-Mining-for-finding-preferences-in-adventure-tourism.pdf). Erişim Tarihi: 16.01.2021.
- SAS. (2020). *Data mining what it is & why it matters*. Erişim Adresi: [https://www.sas.com/tr\\_tr/insights/analytics/data-mining.html#dmusers](https://www.sas.com/tr_tr/insights/analytics/data-mining.html#dmusers). Erişim Tarihi: 15.02.2021.
- SEAGATE. (2020). *Rethink data: Put more of your business data to work from edge to cloud*. Erişim Adresi: [https://www.seagate.com/files/www-content/our-story/rethink-data/files/Rethink\\_Data\\_Report\\_2020.pdf](https://www.seagate.com/files/www-content/our-story/rethink-data/files/Rethink_Data_Report_2020.pdf). Erişim Tarihi: 26.12.2020.
- Shapoval, V., Wang, M. C., Hara, T., & Shioya, H. (2018). Data mining in tourism data analysis: Inbound visitors to japan. *Journal of Travel Research*, 57(3), 310-323. doi:10.1177/0047287517696960
- Sheng, F., Zhang, Y., Shi, C., Qiu, M., & Yao, S. (2020). Xi'an tourism destination image analysis via deep learning. *Journal of Ambient Intelligence and Humanized Computing*, 1-10. doi:10.1007/s12652-020-02344-w

- Siddiqa, A., Karim, A., & Gani, A. (2017). Big data storage technologies: A survey. *Frontiers of Information Technology & Electronic Engineering*, 18(8), 1040-1070. doi:10.1631/Fitee.1500441
- Strohbach, M., Daubert, J., Ravkin, H., & Lischka, M. (2016). Big data storage. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New horizons for a data-driven economy* (pp. 119-141): Springer, Cham.
- Tichter, G., Grossmann, W., & Werthner, H. (1999). Using data mining in analysing local tourism patterns. In *Information and communication technologies in tourism 1999* (pp. 1-10): Springer.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28. doi:10.1257/jep.28.2.3
- Wilkinson, J. S. (2018). Internet of things (iot). In A. E. Grant & J. H. Meadows (Eds.), *Communication technology update and fundamentals* (16 ed., pp. 159-168). New York: Routledge.
- Wolfewicz, A. (2020). *Deep learning vs. Machine learning – what’s the difference?* Erişim Adresi: <https://www.levity.ai/blog/difference-machine-learning-deep-learning>. Erişim Tarihi: 01.14.2021.
- Woodie, A. (2020a). *Data boom gets even bigger, thanks to covid-19*. Erişim Adresi: <https://www.datanami.com/2020/05/07/the-big-cloud-data-boom-gets-even-bigger-thanks-to-covid-19/>. Erişim Tarihi: 29.12.2020.
- Woodie, A. (2020b). *Storage in the exabyte era*. Erişim Adresi: <https://www.datanami.com/2020/02/19/storage-in-the-exabyte-era/>. Erişim Tarihi: 29.12.2020.
- Wu, J., Wang, J., Nicholas, S., Maitland, E., & Fan, Q. (2020). Application of big data technology for covid-19 prevention and control in china: Lessons and recommendations. *Journal of Medical Internet Research*, 22(10), e21980. doi:10.2196/21980
- Xu, J. (2014). *Managing digital enterprise: Ten essential topics*, Springer.
- Yan, Y., Chen, J., & Wang, Z. (2020). Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations. *Applied Geography*, 123, 102306. doi:ARTN 10230610.1016/j.apgeog.2020.102306
- Zhang, J., & Dong, L. (2021). Image monitoring and management of hot tourism destination based on data mining technology in big data environment. *Microprocessors and Microsystems*, 80, 103515.
- Zheng, Z., Zhu, J., & Lyu, M. R. (2013). *Service-generated big data and big data-as-a-service: An overview*. Paper presented at the 2013 IEEE international congress on Big Data.
- Zhu, Y. (2018). A history of communication technology. In A. E. Grant & J. H. Meadows (Eds.), *Communication technology update and fundamentals* (16 ed., pp. 9-21). New York: Routledge.

## **Data Mining and Data Mining Studies in Tourism**

**Tuğba ŞEN KÜPELİ**

Ankara Hacı Bayram Veli University, Faculty of Tourism, Ankara/Turkey

**Kurban ÜNLÜÖNEN**

Ankara Hacı Bayram Veli University, Faculty of Tourism, Ankara/Turkey

### **Extensive Summary**

The progress of information and communication technology and the spread of the internet have made it possible to produce large volumes high speed and diversity of digital data on a global scale, as well as to develop big data applications to store, process, analyze and manage the data. (Akıncı, 2019, p. ii; Ersöz, 2019, p. 13). Modern computer systems can collect very large and diverse data from many different sources, from ATM's to satellites in space. (Bramer, 2007, p. 2). According to the calculation made by the international data center, 1.2 (1.2 trillion gigabytes) of digital data were produced in 2010, with an increase of 0.8 zettabytes compared to the previous year (Press, 2020). In addition, due to the increase in the rate of digitalization due to working from home and education after Covid-19 in 2020, the amount of digital data, which was 45 zettabytes in 2019, surpassed the projections by reaching 59 zettabytes. The amount of global data to be generated by 2025 is estimated from various sources to be 175 or 200 zettabytes (Holst, 2020; Morgan, 2020; Press, 2020; Reinsel, Gantz, & Rydning, 2018). Although the production of large amounts of data dates back to the 1960s, big data started to attract attention with the development of storage and analysis technologies (Jeffcock, 2018; Mayer-Schönberger & Cukier, 2013, p. 14). As a matter of fact, the first step in making data production meaningful is storage. According to Xu (2014, p. 33), storage costs decreased by 33 percent annually between 1990 and 2013 with the advancement of technology, and this technology is advancing rapidly. Unlike hard disk drives, digital storage technologies (for example: NoSQL) enable the development of scalable, reliable, manageable, efficient and flexible data models as well as storing data. Many storage technologies have been developed for different design purposes and benefits such as Google File System, (GFS), Hadoop Distributed File System (HDFS), BigTable, HBase and Scalaris (Siddiqa, Karim, & Gani, 2017). This cloud-based systems try to attract the attention of businesses by promising their users to store, process and analyze data in a more scalable way with affordable and flexible business models (GoogleCloud, 2020). Thus, many industries can be affected by the technological progress and cost-effectiveness in which big data is produced and processed. One of these industries is tourism. Tourists share their experiences on many different media (social media, blogs, booking sites or comment sites, etc.) and shape their vacation plans according to these experiences (Buhalis, 2019; Law, Buhalis, & Cobanoglu, 2014; Mariani, 2019, p. 299; Mistilis & Dwyer, 2000, p. 58-59). Businesses are also integrated into systems that will lead the production of big data in digital media. With data analysis, on the other hand, data can only become a tool that plays an active role in decision-making processes. Big data analytics come into play at this point, triggering the development of digital technologies such as artificial intelligence, data mining and deep learning, and it can enable to learn past and current behaviors of consumer and determine the needs of tourists (Mariani, 2019, p. 301). With the increasing importance of data analysis, it has become critical to be able to obtain the information hidden within the data. For example, it can be used in an organization to continue its operations, to make significant discoveries in various fields, to define the causes and treatments of different diseases, or to forecast weather and

natural disasters (Bramer, 2007, p. 2). For this reason, various types of data are stored by different public institutions, businesses and other organisations, depending on their purposes (Dolgun, Özdemir, & Oğuz, 2009, p. 48). With data mining techniques, certain conclusions are drawn and summarized by pursuing the traces in the data (Mayer-Schönberger & Cukier, 2013, p. 16). As a matter of fact, the transformation of big data into an opportunity depends on its ability to process and discover the information it contains, adapting this information to business processes or producing a specific solution depending on the industry.

When the large volume of data, produced with the effect of digitalization, is analyzed correctly, it has the potential to offer various opportunities in terms of sustainable and innovative tourism planning and marketing and eventually to increase the competitiveness of the destination (Park et al., 2020, p.14-15). So that the intention to return to a destination depends on the satisfaction level of the visitors with their goods and services (Shapoval et. al., 2018, p. 23), and to achieve this result, central and local governments and touristic enterprises in particular (hotels, restaurants, airlines, travel agencies, etc.) are trying to develop various strategies. Among these, data mining is emerging as a concept that has an increasing use area. Data mining techniques enable strategic marketing planning (Mirela, Tiberiu, & Adina, 2009), help to understand tourist behavior (Tichter, Grossmann, and Werthner, 1999) and manage to customer relationship (Olmeda & Sheldon, 2002), can drive effective demand forecasting (Law, Mok and Goh, 2007) and can provide a more comprehensive analysis opportunity than classical statistical analyzes on issues such as identification of problems regarding destinations (Gu et. al., 2016), repeat visit intention of tourists (Shapoval et. al., 2018), analysis of tourist profile, estimation of tourist number and tourist expenditure (Bose, 2009, p. 936).

Although the studies in the field of tourism are increasing gradually, since the data mining issue has been described as new concept, there is still a need for studies to be done as in other fields. The theoretical development of the concept in the tourism literature is also lacking due to the fact that the studies dealing with the issue of data mining in terms of its technical aspects are predominant in the literature and the data, belongs to tourism industry, is mostly treated as a "means (sample data)". Thus, there arises the need to comprehensively address the concept of data mining in the context of tourism (Bach et al., 2013; Li et al., 2019). Based on this requirement, in the present review, the applications in the current tourism literature are discussed along with concepts such as big data and artificial intelligence by describing the principle of data mining. In this way, a general perspective is presented on solutions for tourism dynamics with data mining techniques. In addition, we examined the distribution of data mining studies on the "Google Academic" between 1999 and 2020, it was reviewed that some studies directly related to the field of tourism brought solutions to different tourism dynamics. In addition, studies on data mining in the field of tourism can be divided into two different categories. Some of the studies are based on large data sets produced in the tourism industry (number of tourists or overnight stays, spending amounts, restaurant reviews, hotel ratings, etc.) and the main purpose is to test the developed algorithm (software) or analysis technique and to calculate success rates. In the other category, the main purpose is to contribute to the tourism industry and the literature, and besides the data set used, the analyzes, comments and evaluations made also include studies related to the field of tourism. The number of studies in the second category is relatively small. This situation is also shown in the study results of Bach, Schatten and Marušić (2013) and Li et al. (2019). Although current studies on data mining related to tourism industry offer various solutions for tourism dynamics, it could be stated that they are not sufficient for the development of the concept. There is a need for studies on data mining, which are analyzed and evaluated especially with tourism perspective and field knowledge (Li et. al., 2019).