



Web Tabanlı Mekânsal İçerikli Dijital Veri Kaynaklarının Değerlendirilmesinde Yeni Yaklaşımlar, Teknolojik Gelişmeler ve Kentsel Uygulama Örnekleri*

New Approaches, Technological Advancements, and Urban Applications in the Evaluation of Web-Based Spatial Digital Data Sources.

Lütfiye KUŞAK¹ , Doğan UÇAR²

Öz

Son yıllardaki teknolojik gelişmeler, web tabanlı mekânsal içerikli dijital verilerin elde edilme, yönetim ve kullanım süreçlerinde önemli dönüşümler yaratmıştır. Bu derleme çalışmasında, web kazıma ve Uygulama Programlama Arayüzü (API) temelli veri çıkarma yöntemleriyle dijital kaynaklardan mekânsal verilerin elde edilmesi ve kullanım alanları değerlendirilmiştir. Literatürde bu tür verilerin kentsel planlama, altyapı ve afet yönetimi gibi alanlarda etkin biçimde kullanıldığı görülmektedir. Web kazıma ve uygulama programlama arayüzü yaklaşımları, geleneksel veri toplama yöntemleri ile elde edilen mekânsal veri setlerini lokasyon ve öznitelik bilgileri açısından destekleyebilmektedir. Ancak veri bütünlüğü, gizlilik ve standardizasyon eksiklikleri; web tabanlı mekânsal içerikli dijital veri kaynaklarının daha etkin bir şekilde kullanılmasının önündeki önemli zorluklardır. Buna rağmen bu dijital kaynaklar, kentsel planlama ve yönetime bütüncül bir bakış kazandırarak tematik veri entegrasyonunu destekleyen güçlü araçlar hâline gelmiştir. Sonuç olarak, web tabanlı mekânsal içerikli dijital veri kaynakları, yalnızca veriye erişimi kolaylaştırmakla kalmayıp; kentsel planlama ve yönetimde farklı tematik verilerin entegrasyonunu sağlayarak daha bütüncül analiz ve stratejilerin geliştirilmesine katkı sunmaktadır.

Anahtar Kelimeler: Mekânsal Veri, Web kazıma, Uygulama Programlama Arayüzü, Lokasyon.

ABSTRACT

Recent technological advances have significantly transformed the acquisition, management, and use of web-based spatial data. This review evaluates the acquisition and application areas of spatial data from digital sources using web scraping and Application Programming Interface (API) based data extraction methods. The literature shows that such data are effectively used in areas such as urban planning, infrastructure management, and disaster management. Web scraping and API approaches can supplement spatial datasets obtained through traditional data collection methods by providing location and attribute information. However, data integrity, privacy, and standardization deficiencies are significant challenges to the more effective use of web-based spatial digital data sources. Despite this, these digital sources have become powerful tools that support thematic data integration, providing a holistic perspective on urban planning and management. Consequently, web-based spatial data sources not only facilitate data access but also support more holistic analyses and strategies by enabling the integration of diverse thematic datasets in urban planning and management.

Keywords: Spatial Data, Web Scraping, Application Programming Interface, Location.

¹ Sorumlu Yazar: Dr. Öğr. Üyesi Lütfiye KUŞAK, Mersin Üniversitesi, lutfyekusak@mersin.edu.tr, ORCID: 0000-0002-7265-245X

² Prof. Dr. Doğan UÇAR, İstanbul Teknik Üniversitesi, ducar49@gmail.com, ORCID: 0009-0003-0553-7704

*Bu makale, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü'nde 2010 yılında Lütfiye KUŞAK tarafından Prof. Dr. Doğan UÇAR danışmanlığında yapılan Web ortamında mekânsal veri kazanımı adlı doktora tezinden üretilmiştir.



GİRİŞ

Bilimsel araştırmalar, kamu politikaları ve teknolojik gelişmelerin temelinde güvenilir veriler bulunmaktadır. Veriler, yalnızca mevcut durumun anlaşılmasını değil; aynı zamanda geleceğe yönelik stratejik planlamaların yapılabilmesini de mümkün kılmaktadır. Dijital teknolojilerin yaygınlaşmasıyla birlikte ölçüm ve gözlem yolu ile elde edilen veriler, dijital kaynaklarla da desteklenmeye başlanmıştır. World Wide Web (WWW) ile birlikte dijital ortamda veri üretimi ve erişimi hız kazanmıştır. Ekim 2025 verilerine göre, dijital dünyada yaklaşık 1,2 milyar web sitesinin bulunduğu ve her gün ortalama 252.000 yeni sitenin dijital dünyaya eklendiği raporlanmıştır (Haan, 2025). Özellikle sosyal medya platformları, haber siteleri ve bloglar gibi kullanıcı etkileşiminin yoğun olduğu dijital ortamlarda, sürekli artan veri üretimi dikkat çekmektedir.

Dijital veri üretimindeki bu hızlı artış, verilerin elde edilmesi, yönetilmesi ve analiz edilmesi konularında web kazıma ve uygulama programlama arayüzü gibi yöntemleri gündeme getirmiştir. Bu yöntemlerle toplanan veriler; pazarlama, şehircilik, sağlık, turizm, çevre yönetimi ve afet analizi gibi farklı disiplinlerde karar destek süreçlerinde önemli rol oynamaktadır. Özellikle kullanıcı etkileşiminin yüksek olduğu veriler, bireylerin ürün, olay veya hizmetlere ilişkin tutumlarını, afet ve kriz dönemlerinde toplumsal davranış kalıplarının anlaşılmasını (Dolu & Şen, 2024) analiz etmek amacıyla kullanılan duygu analizi çalışmalarında temel bir girdi sağlamaktadır.

2000’li yıllardan itibaren coğrafi bilgiye erişim ve mekânsal veri analitiği alanlarında yürütülen araştırmalar ile hem teorik hem de uygulamalı düzeyde önemli ilerlemeler kaydedilmiştir. Bu süreçte web tabanlı dijital kaynaklardan, mekânsal içerikli verilerin toplanması ve yeniden kullanılabilir biçimlerde sunulması çalışmaları hız kazanmıştır (Martins, 2008).

Mekânsal içerikli dijital veri kaynaklarının artan erişilebilirliği, kentlerin analiz, planlama ve yönetiminde de yeni fırsatlar yaratmaktadır. Uzaktan algılama verileri, açık veri portalları (Özbek & Ercoşkun, 2023) ve sosyal medya platformları aracılığıyla şehirlerin fiziksel yapısı, ulaşım ağları ve çevresel riskleri hakkında çok boyutlu bilgiler kısa sürede elde edilebilmektedir. Bu çok kaynaklı veri setleri, afet risklerinin izlenmesi (Memduhoğlu, 2024), ulaşım ağlarının optimizasyonu (Molina-Campoverde vd., 2024) ve kentsel büyüme dinamiklerinin analizi gibi geniş bir uygulama yelpazesinde kullanılmaktadır. Kamu kurumları ve yerel yönetimler ise bu verileri karar destek sistemleri, politika üretimi ve katılımcı planlama süreçlerine entegre ederek, kent yönetiminde veri odaklı yaklaşımları güçlendirmektedir (Taşkanat vd., 2018).

Sonuç olarak mevcut çalışmalar, mekânsal içerikli dijital verilerin artan erişilebilirliğinin kentlerin analiz, planlama ve yönetim süreçlerinde veri odaklı, çok boyutlu yaklaşımların gelişimine büyük katkı sağladığını göstermektedir. Bu çalışmada, kentsel analizlerde kullanılan dijital veri kaynakları, bu verilerin elde edilmesinde kullanılan teknikler ve veri madenciliği yaklaşımları değerlendirilmiştir. Ayrıca, Coğrafi Bilgi Sistemleri (CBS) ile entegrasyon ve görselleştirme çalışmalarına değinilmiş, farklı kent çalışmalarına ait uygulama örnekleri sunulmuş; karşılaşılan zorluklar ve gelecekteki araştırma yönelimleri tartışılmıştır.

1. Dijital Veri Kaynaklarının Tarihsel Gelişimi

1990’lı yıllardan itibaren internet ve web teknolojilerindeki hızlı gelişmeler, dijital verinin üretilme, paylaşılma ve erişim biçimlerinde köklü bir dönüşüm başlatmıştır. Bu tarihsel dönüşümün incelenmesi, günümüzde kullanılan dijital veri çıkarım tekniklerinin ve şehircilik alanındaki uygulamaların anlaşılması açısından önemli bir temel sunmaktadır.

1.1 Web Tabanlı Dijital Veri Çıkarım Süreçlerinin Evrimi

Web tabanlı veri kaynaklarının çeşitlenmesi ve kullanıcı erişiminin kolaylaşması, veri çıkarım yöntemlerinin de otomasyon, yazılım ve algoritma temelli yaklaşımlar doğrultusunda yeniden şekillenmesine yol açmıştır (Şekil 1).



Şekil 1. Veri Çıkarım Teknolojilerinin Tarihsel Gelişimi. Şekil çalışmalardan ve kitaplardan sentezlenmiştir (Laender vd., 2002; Makice, 2009; Mitchell, 2018; Mueller, 2004; Petroutsos, 2014; Richardson, 2025; Scrapy, 2025).

1990'lı yılların başlarında dijital veri çıkarım süreçleri henüz gelişmemiş olduğundan, veriler çoğunlukla manuel yöntemlerle elde edilmekteydi. Kullanıcılar, web içeriklerini elle kopyalayarak veya basit betikler aracılığıyla veri çekme işlemleri uygulayarak gerçekleştirmekteydiler (Laender vd., 2002). 2000'li yıllarda ise Perl ve Python dilleriyle geliştirilen komut dosyaları yaygınlaşmıştır. Geliştirilen BeautifulSoup kütüphanesi HTML belgelerinin ayrıştırılmasını kolaylaştırmıştır (Richardson, 2025). Ardından geliştirilen Scrapy, büyük ölçekli veri kazımlarında kullanımı yaygınlaşmıştır (Scrapy, 2025). Aynı dönemlerde Amazon API (Mueller, 2004), Twitter API (Makice, 2009) ve Google Maps API (Petroutsos, 2014) gibi arayüzler, geliştiricilere gerçek zamanlı veri akışı sağlamıştır. 2015 sonrası dönemde ise web kazıma süreçleri makine öğrenimi ve doğal dil işleme yöntemleriyle desteklenmiştir. Octoparse, ParseHub ve Diffbot gibi kullanıcı dostu araçlar teknik bilgi gerektirmeden veri çıkarımını mümkün kılmıştır. Web tarayıcı botlarının yaygınlaşmasıyla etik ve yasal sorunlar da artmıştır. Bu kapsamda, otomatik yazılımların hangi sayfalara erişebileceğini belirten robots.txt dosyaları (Mitchell, 2018) ve site kullanım koşulları rehber niteliği kazanmıştır. Ayrıca veri gizliliği düzenlemelerinin yanı sıra, kişisel verilerin toplanması ve işlenmesi konusunu netleştirmek için Türkiye'de Kişisel Verilerin Korunması Kanunu (KVKK), uluslararası ölçekte California Civil Code (CCPA) gibi çeşitli yasal normlar geliştirilmiştir.

1.2 Mekânsal Özellik İçeren Dijital Veri Kaynaklarının Gelişimi

Mekânsal verinin dijital ortamlarda üretimi, paylaşımı ve kullanımı zaman içinde önemli teknolojik gelişmelerle dönüşüme uğramıştır (Şekil 2).



Şekil 2. Mekânsal Özellik İçeren Dijital Veri Kaynaklarının Tarihsel Gelişimi. Şekil, uzaktan algılama, CBS yazılımları, web tabanlı platformlar, büyük veri, yapay zekâ ve nesnelerin interneti teknolojileri üzerine yapılan çalışmalardan ve kitaplardan sentezlenmiştir (Burrough & McDonnell, 1998; Goodchild, 2007; Petroutsos, 2014; Tomlison, 1974).

1960–1980 yılları arasında ilk sayısal harita üretimleri ve analog verilerin dijital ortama aktarımıyla dijital mekânsal veri üretiminin temelleri atılmıştır (Burrough & McDonnell, 1998). 1972’de Landsat-1 uydusunun fırlatılmasıyla uzaktan algılama yöntemleri yaygınlaşmış, ayrıca CBS uygulamalarının ilk örnekleri sunulmuştur (Tomlison, 1974). 1980–1995 döneminde GPS’in sivil kullanıma açılması, konumsal doğrulukta devrim yaratmış; kurumlar kent planlama ve çevre yönetiminde dijital sistemleri benimsemiştir. Topolojik veri modelleri ve Sayısal Yükseklik Modeli (SYM) formatları bu dönemde geliştirilmiştir. 1995–2005 arasında internetin yaygınlaşmasıyla mekânsal veriler web ortamında paylaşılabılır hâle gelmiş, MapQuest ve Yahoo Maps gibi ilk web tabanlı harita servisleri kullanıma sunulmuştur. 2005–2015’te Gönüllü Coğrafi Bilgi (VGI) kavramı önem kazanmış; OpenStreetMap, Google Earth ve Wikimapia gibi platformlar halkı veri üreticisi konumuna getirmiştir (Goodchild, 2007). 2015 sonrası dönemde nesnelerin interneti, sensör teknolojileri, sosyal medya ve mobil cihazlar aracılığıyla gerçek zamanlı mekânsal veri üretimi artmış; bu veriler CBS, makine öğrenmesi ve yapay zekâ algoritmalarıyla analiz edilmeye başlanmıştır.

2. Kentsel Çalışmalar için Dijital Veri Kaynakları ve Elde Edilme Yöntemleri

Şehircilik alanında yürütülen çalışmalarda, geleneksel ve dijital veri kaynaklarının ayrı ayrı ya da birlikte kullanıldığı pek çok örnek bulunmaktadır. Bu kaynakların tür, kapsam, doğruluk, güncellik, erişilebilirlik gibi temel özelliklerinin incelenmesi, şehircilik çalışmalarında veri seçimi ve kullanım biçimlerinin anlaşılmasına katkı sağlayacaktır.

2.1 Veri Kaynakları ve Özellikleri

Geleneksel veri kaynakları, tarihsel sürekliliği olan, tanımlı yöntemlerle üretilen ve yetkili kurumlarca doğrulanmış güvenilir bilgileri ifade eder. Bu veriler, standart formatlarda kayıt altına alınır ve belirli yasal çerçevelere dayanır. Nüfus kayıtları, arazi kullanımı, sınırlar, gayrimenkul değerleri, haritalar, altyapı ve ulaşım bilgileri gibi veriler geleneksel veri kaynakları arasında yer almaktadır (Goodchild, 2007). Güncelleme sıklıkları genellikle belirli periyotlara dayanır. Doğrudan gözlem, kontrol ve standardizasyon mekanizmaları doğruluğu artırsa da manuel girişler veya yorum hataları veri kalitesini etkileyebilir. Ayrıca, personel ve zaman gereksinimleri nedeniyle bu tür veri üretimi, yüksek maliyetli olabilmekte ve kurumsal erişim izinleri, hükümet politikaları da verilerin paylaşımını sınırlandırabilmektedir.

Dijital veri kaynakları ise, teknolojik gelişmelerle birlikte kentsel dinamikleri çok boyutlu biçimde analiz etmeye olanak tanıyan geniş ve çeşitli veri türlerini kapsamaktadır. Twitter, Instagram, Facebook gibi sosyal medya platformları, resmî kurum siteleri, emlak portalları, TripAdvisor gibi ticari platformlar ve Gönüllü Coğrafi Bilgi (VGI) sistemleri web tabanlı veri kaynakları arasında yer almaktadır. Bu kaynaklardan elde edilen veriler, çevresel değişkenlerin ve kentsel sistemlerin gerçek zamanlı izlenmesine imkân tanımaktadır. Web tabanlı veriler, sağlayıcının ölçeğine bağlı olarak küresel veya yerel karakteristikler gösterebilir. OpenStreetMap, sosyal medya ve uydu görüntüleri küresel veri kaynaklarına; belediyelerin açık veri portalları ve şehir sensörleri ise yerel dijital kaynaklara örnektir. Bu veriler genellikle otomatik sistemler aracılığıyla gerçek zamanlı veya düzenli aralıklarla güncellenmekte, GPS entegrasyonu sayesinde yüksek konumsal doğruluğu desteklenebilmektedir. Ancak kullanıcı katkısına dayalı içerikler veya doğrulama standartlarının bulunmadığı platformlarda hatalar bulunabilmektedir. Maliyet açısından değerlendirildiğinde, gerekli altyapı kurulduktan sonra dijital veri kaynakları genellikle düşük maliyetli olmakla birlikte, veri gizliliği, doğruluk kontrolünün eksikliği, kullanıcı katkısının denetlenememesi ve teknolojik eşitsizlikler bu tip veri kaynaklarının kullanımında dikkat edilmesi gerekli durumlar arasında yer almaktadır. KVKK, web kazıma sınırlamaları ve sınır aşan veri aktarımı düzenlemeleri veri çıkarım süreçlerini doğrudan etkilemektedir. Ayrıca, telif hakları ve Creative Commons, Open Database License gibi açık veri lisansları veri paylaşım koşullarını belirlemektedir.

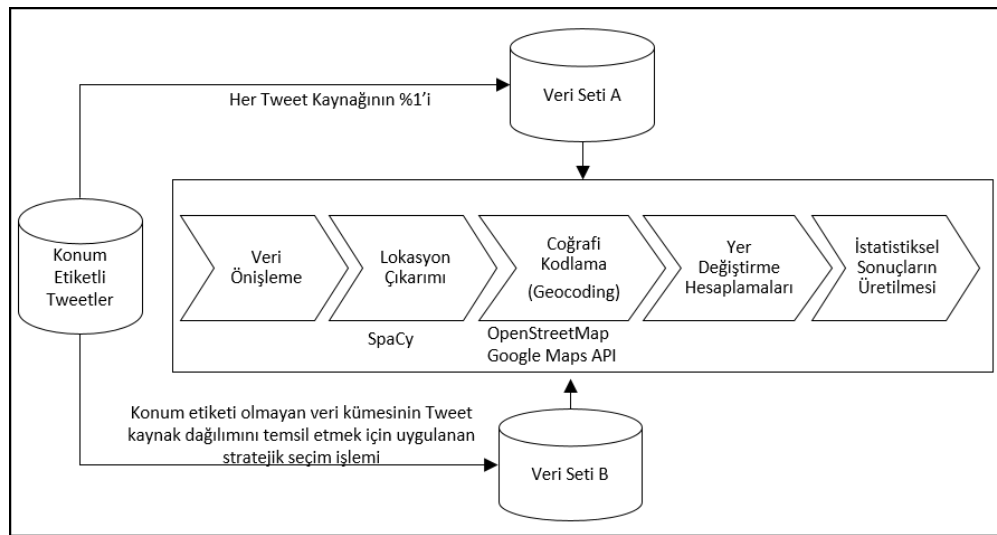

```

Start:
# Initialize the geolocator
Initialize Geolocator with a user agent
# Geocoding (Find coordinates by address)
Location = Geolocator.geocode(Address) # Convert address to coordinates
If Location is found:
    Print "Address: " + Location.address # Print the address
    Print "Coordinates: " + (Location.latitude, Location.longitude) # Print the coordinates
Else:
    Print "Address not found."
# Reverse Geocoding (Find address by coordinates)
ReverseLocation = Geolocator.reverse(Coordinates) # Convert coordinates to address
If ReverseLocation is found:
    Print "Address: " + ReverseLocation.address # Print the address for the given coordinates
Else:
    Print "Address not found."
End.

```

Şekil 4. Coğrafi Kodlama Süreçlerine İlişkin Örnek Betik Akışı.

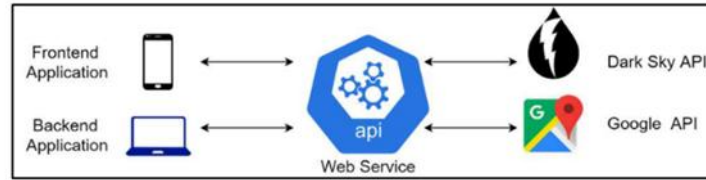
Coğrafi kodlama süreçleri, yalnızca web kazıma verileriyle sınırlı olmayıp, farklı dijital veri kaynaklarının mekânsal analiz süreçlerinde yaygın biçimde kullanılmaktadır. Mevcut adreslerin veya web kazıma işlemleri sonucunda elde edilen adreslerin koordinatlara dönüştürülmesi için coğrafi kodlama, uygulama programlama arayüzü veya sensör tabanlı verilerde ise koordinatlardan adres bilgisinin elde edilmesi için ters coğrafi kodlama tercih edilmektedir (Şekil 5). Bu süreçlerde Google Maps Geocoding API, OpenStreetMap tabanlı Nominatim, Mapbox Geocoding API ve GeoNames gibi çevrimiçi servisler yaygın olarak kullanılmaktadır (Serere vd., 2023). Ayrıca PostGIS fonksiyonları, ArcGIS Geocoding Service ve QGIS Geocode plugin gibi CBS tabanlı araçlar da bu süreçte tercih edilebileceği gibi Python ortamında Geopy, Googlemaps ve OSMnx kütüphaneleri kullanılarak bu işlemler otomatikleştirilebilmektedir.



Şekil 5. Elde Edilen Lokasyonların Koordinatlandırılmasında OpenStreetMap's Nominatim ve Google Maps API Kullanımı (Serere vd., 2023), tarafından sunulan şekil temel alınarak Türkçe'ye çevrilmiştir.

Uygulama programlama arayüzleri, farklı yazılım bileşenlerinin veri ve hizmet paylaşımını sağlayan iletişim mekanizmalarıdır. Büyük veri platformlarının üçüncü taraf geliştiricilere açık veri sunmasını sağlayan uygulama programlama arayüzleri, JSON, XML veya CSV gibi yapılandırılmış formatlarda veri sağlayarak web kazıma kıyasla daha standart, güvenilir ve pratik bir yöntem sunar. Farklı tematik alanlarda veri sunan uygulama programlama arayüzleri, meteorolojik ve çevresel gözlemler, konum ve harita servisleri, afet ve jeolojik bilgiler, ulaşım altyapısı ile sosyal medya tabanlı veriler gibi çok

çeşitli kaynaklardan mekânsal veri elde edilmesine olanak sağlamaktadır. OpenWeatherMap, NOAA, Copernicus, Dark Sky gibi servisler meteorolojik (Şekil 6) (Setiawan vd., 2023) ve çevresel veriler sağlarken; Google Maps, OpenStreetMap, Mapbox ve Overpass API'leri konum, adres ve ilgi noktası (POI) tabanlı mekânsal veriler üretmektedir (Kilic & Gülgen, 2020; Serere vd., 2023). Google Directions API'leri ulaşım bilgilerine; Twitter, Foursquare ve Facebook ise sosyal medya etkileşimlerine ilişkin konum tabanlı içeriklere erişim sunmaktadır. Uygulama programlama arayüzlerinin tümü doğrudan mekânsal veri sağlamamakla birlikte, çoğu coğrafi referans içeren veya mekânsal analizlere entegre edilebilen tematik veriler üretmekte ve bu yönüyle mekânsal karar destek sistemlerinin geliştirilmesine önemli katkılar sağlamaktadır.



Şekil 6. Hava Tahmini ve Coğrafi Konum Teknolojisine Dayalı Seyahat Rota Öneri Sistemi için Uygulama Programlama Arayüzü Kullanımı (Setiawan vd., 2023).

Sonuç olarak, uygulama programlama arayüzleri, güvenilir, yapılandırılmış ve standartlaştırılmış veri erişimi sağlamaları bakımından web kazımaya göre önemli avantajlar sunmaktadır. Ancak günlük istek kotası, çağrı başına veri sınırı ve belirli veri alanlarına erişim kısıtlamaları gibi sınırlamalar bulunmaktadır. Ayrıca coğrafi etiketlerin eksik veya hatalı olması konum doğruluğunu etkileyebilmektedir. Farklı uygulama programlama arayüz sağlayıcılarının sunduğu coğrafi kodlama hizmetleri doğruluk ve ayrıntı düzeyi açısından değişiklik gösterebilir (Serere vd., 2023).

2.3 Analiz ve Görselleştirme İşlemleri

Dijital veri kaynaklarından elde edilen bilgilerin analizinde veri madenciliği teknikleri önemli bir rol oynamaktadır. Bu yöntemler, büyük ve karmaşık veri kümelerinden anlamlı, yeni ve yararlı bilgilerin çıkarılmasını amaçlar (Han vd., 2012). Günümüzde web siteleri, sosyal medya içerikleri, sensör kayıtları, uydu görüntüleri, GPS verileri ve dijitalleştirilmiş belgeler gibi çok sayıda kaynaktan veri üretilmektedir. Bu verilerin anlamlandırılması için ise veri madenciliği süreçlerinden yararlanılmaktadır. Veriler genellikle yapılandırılmış, yarı yapılandırılmış veya yapılandırılmamış biçimlerde sınıflandırılır ve uygulanacak analiz teknikleri bu türlere göre belirlenir. Yapılandırılmış verilerde geleneksel veri madenciliği, yarı yapılandırılmış web verilerinde web madenciliği, yapılandırılmamış verilerde ise metin madenciliği yöntemleri kullanılmaktadır.

Web madenciliği, internet üzerindeki yapısal veya yarı yapısal verilerin analizine odaklanır ve web sayfalarının HTML yapısı, bağlantı ilişkileri, kullanıcı etkileşimleri ve gezinme davranışlarını temel veri kaynağı olarak değerlendirir. Web madenciliğinin alt disiplini olan web içerik madenciliği, HTML sayfaları, blog yazıları ve haber siteleri gibi kaynaklardan bilgi çıkarmayı amaçlar ve metin dışı öğeleri de analiz eder (Rajnish & Srivastava, 2023). Buna karşılık metin madenciliği yalnızca yazılı içeriğe odaklanır ve özellikle sosyal medya, bloglar ve haber metinleri gibi yapılandırılmamış verilerin analizinde etkilidir (Makhabel vd., 2017). Bu süreçlerde doğal dil işleme teknikleri, metinlerden anlamlı bilgi elde etmek amacıyla kullanılmaktadır (Jo, 2024).

Veri madenciliği yaklaşımları, veri yapısına göre farklılık göstermektedir. Yapılandırılmış verilerde sınıflandırma, kümeleme veya ilişki kuralı madenciliği gibi istatistiksel ve makine öğrenmesi tabanlı K-ortalamlar, karar ağaçları, Apriori gibi yöntemler uygulanırken; yarı yapılandırılmış web verilerinde bağlantı, içerik ve kullanıcı davranışı analizine dayalı web madenciliği öne çıkar. Yapılandırılmamış verilerde ise metin madenciliği ve doğal dil işleme yöntemleri kullanılarak sosyal medya, haber ve

adres verilerinden anlamlı örüntüler elde edilmektedir. Çalışmalarda TF-IDF, Word2Vec, Levenshtein, Jaro-Winkler, kosinüs benzerliği gibi yaklaşımlar tercih edilmektedir.

Bu yöntemlerin yerel ölçekteki uygulamalarına örnek olarak, Şişli ilçesi için yürütülen bir çalışmada web spider kullanılarak adres ve bina kullanım fonksiyonu verileri toplanmış; bu veriler üzerinde TF-IDF, Öklid mesafesi ve Kosinüs benzerlik analizleri uygulanmıştır. Elde edilen sonuçlar, İstanbul Büyükşehir Belediyesi Adres Bilgi Sistemi ile karşılaştırılmış ve yöntemin eksik veya güncellenmemiş kayıtları tamamlayıcı nitelikte olduğu belirlenmiştir (Kuşak, 2010). Bu tür uygulamalar, dijital veri kaynaklarının kentsel ölçekte mekânsal doğrulama ve veri bütünlüğü analizlerinde etkili biçimde kullanılabileceğini göstermektedir.

Benzer biçimde, Chow vd. (2018) Teksas'ta yaşayan Vietnam kökenli bireylerin göç örüntülerini analiz etmiş; web kazıma yoluyla elde edilen adres verilerini Google Maps API aracılığıyla doğrulamış ve veri madenciliği teknikleriyle göç hareketlerini sınıflandırmıştır. Elde edilen veriler CBS ortamında işlenerek göç yolları, yoğunluk haritaları ve demografik etkileşimler ile görselleştirilmiştir. Bu çalışma ise, web tabanlı verilerin CBS tabanlı görselleştirme ve mekânsal yorumlama potansiyelini göstermektedir.

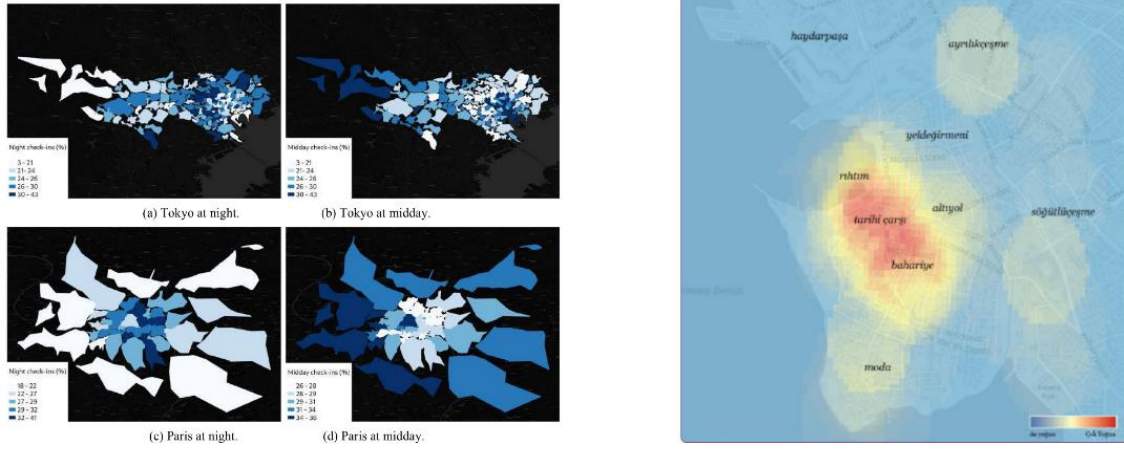
Mekânsal veri madenciliği, klasik veri madenciliğinden farklı olarak konum bilgisi, mekânsal otokorelasyon ve topolojik ilişkileri dikkate alır. Bu yaklaşım, afet yönetimi, kentsel planlama, çevresel analizler, ulaşım optimizasyonu ve halk sağlığı gibi fiziksel mekânla ilişkili alanlarda kullanılmaktadır (Li vd., 2015). Coğrafi ilişkilerin ve mekânsal desenlerin belirlenmesinde CBS desteğiyle uygulanan bu teknikler kritik öneme sahiptir.

Veri madenciliği algoritmalarının güvenilir sonuçlar üretebilmesi, veri setlerinin doğruluk, tutarlılık ve güncellik düzeyine bağlıdır. Büyük veri setlerinde biçimsel uyumsuzluklar, kalite farklılıkları ve güncelleme eksiklikleri genellikle karşılaşılan sorunlar arasında yer almaktadır (Brindley vd., 2018). Bu sorunlar, eksik verilerin tahmine dayalı yöntemlerle tamamlanması, hatalı değerlerin temizlenmesi, değişkenlerin normalizasyonu ve veri entegrasyonu gibi işlemlerle giderilebilir. Ayrıca kayıt eşleştirme ve zaman serisi modelleme gibi tekniklerle veri bütünlüğü artırılabilir.

Sonuç olarak, veri madenciliği yaklaşımları büyük ölçekli, çok kaynaklı ve karmaşık dijital verilerin analizine imkân tanıyarak kentsel planlama, afet yönetimi, ulaşım ve çevre yönetimi gibi alanlarda güçlü bir karar destek aracı sunmaktadır. Dijital verilerin CBS ile entegrasyonu ise, mekânsal örüntülerin belirlenmesi ve verilerin görsel olarak yorumlanabilmesi açısından önemli avantajlar sağlamaktadır.

3. Web Tabanlı Dijital Veri Kaynaklarının Kentsel Analiz ve Uygulamalarda Kullanımı

Günümüzde web sayfaları, sosyal medya içerikleri, uydu görüntüleri, sensör verileri ve açık veri platformları gibi çeşitli kaynaklardan elde edilen veriler; CBS, veri madenciliği ve kentsel modelleme teknikleriyle analiz edilerek mekânsal örüntülerin ve eğilimlerin belirlenmesinde kullanılmaktadır. Bu yaklaşımlar, veri temelli şehir planlaması, afet yönetimi, ulaşım ve çevre yönetimi gibi alanlarda karar destek sistemlerinin geliştirilmesine katkı sağlamaktadır. Büyük veri setlerinin şehircilikteki yaygın kullanımı, kentsel mekânın fiziksel yapısı ile sosyal davranış örüntülerinin çok boyutlu analizini desteklemektedir. Mobil cihazlar, sosyal medya ve uzaktan algılama verileri, kentlerin dinamik yapısını bütüncül biçimde değerlendirmeyi sağlar. Bu veriler, trafik, nüfus hareketliliği ve sosyal etkileşimlerin analizinde (Şekil 7a) güçlü bir temel oluşturarak sürdürülebilir yaşam alanlarının tasarımına katkı sunmaktadır (Calafiore vd., 2021). Ayrıca, sosyal medya verilerinin (Şekil 7b) mekânsal analizlerle bütünleştirilmesi yenilikçi şehircilik yaklaşımlarını desteklemektedir (Üsküplü & Çolakoğlu, 2019).

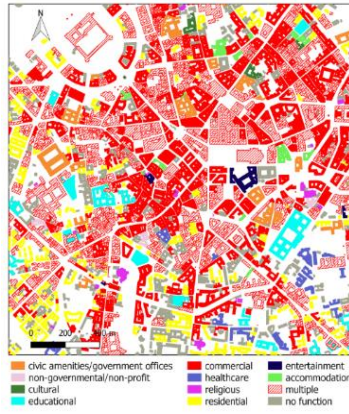


- a. Zaman bakımından benzer coğrafyaya sahip iki örnek şehir ve Foursquare mekânlarıyla insan dinamiklerinin işlevsel ve bağlamsal analizi (Calafiore vd., 2021)
- b. Sosyal Ağ Verileri ve Mekân Dizim Analizlerinin Kentsel Stratejiler Geliştirmede Kullanımı: Kadıköy Örneği (Üsküplü & Çolakoğlu, 2019)

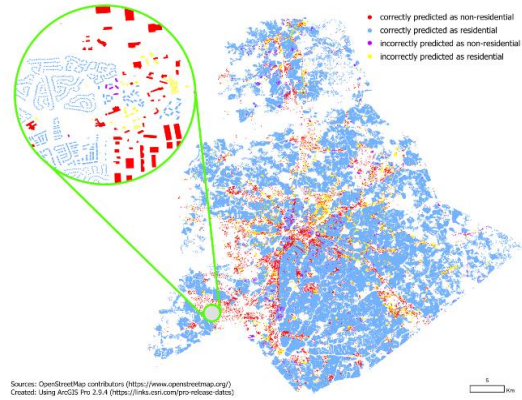
Şekil 7. Dijital veri kaynaklarının şehircilik çalışmalarında kullanım örnekleri

Kentsel analizlerde kullanılan dijital veriler, kaynak ve yapı bakımından büyük çeşitlilik göstermektedir. OSM gibi gönüllü coğrafi bilgi platformları, konum tabanlı analizlerde yaygın biçimde değerlendirilen açık veri kaynakları arasındadır. Özellikle ilgi noktası verileri, restoranlar, parklar ve kültürel alanlar gibi mekânların mekânsal örüntülerinin belirlenmesinde önemli rol oynamaktadır. Wang vd. (2016), Pekin kent merkezinde dinamik arazi kullanım desenlerini belirlemek amacıyla Sina Weibo verileriyle ticari POI verilerini birleştirmiş; K-ortalamalar ve Word2Vec yöntemleriyle 9,5 milyon sosyal medya iletisi ve 385.000 POI kaydını analiz ederek konut, ticaret ve ulaşım merkezlerini tespit etmiştir. Google Places API, Foursquare ve OSM (Küçük & Anbaroğlu, 2019) gibi servisler lokasyon bazlı veri toplamada; sosyal medya içerikleri ise kullanıcı hareketliliği ve mekânsal-zamansal dinamiklerin analizinde etkin biçimde kullanılmaktadır.

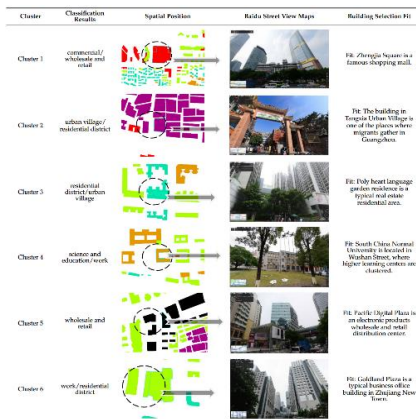
Bina verileri, lokasyon temelli analizlerin bileşenlerinden birisidir ve resmî veri tabanları, açık veri portalları, emlak siteleri veya web kazıma yöntemleriyle elde edilebilir. Bina fonksiyonu, bir yapının toplumsal veya ekonomik amaçla kullanımını tanımlayan önemli bir sınıflandırma olup planlama, altyapı ve sürdürülebilirlik analizlerinde kritik öneme sahiptir. Konut, ticaret, sanayi, kamu, eğitim, sağlık ve eğlence gibi kategoriler bu kapsamda değerlendirilir (Vaggione, 2014). CBS destekli yaklaşımlar (Şekil 8a), sahip olduğu dijital veri entegrasyonu sayesinde daha doğru ve güncel mekânsal analizler yapılmasına olanak tanımaktadır (Fonte vd., 2018). Atwal vd. (2022), OSM öznetelikleri ve mekânsal ilişkileri kullanarak makine öğrenmesi ile binaları konut ve konut dışı olarak sınıflandırmıştır (Şekil 8b). Gao vd. (2021), Tencent kullanıcı yoğunluğu verilerini kullanarak %83'ün üzerinde doğrulukla bina türlerini belirlemiştir (Şekil 8c). Farklı veri kaynaklarının bütünleştirilmesi, bina fonksiyonlarının sınıflandırma doğruluğunu ve kapsayıcılığı artırmaktadır. Uydu görüntüleri, Gaode Map (Şekil 8d) ve saha anketleri gibi yöntemlerin birlikte kullanıldığı çalışmada sınıflandırma başarısının yükseldiği görülmektedir (Deng vd., 2022).



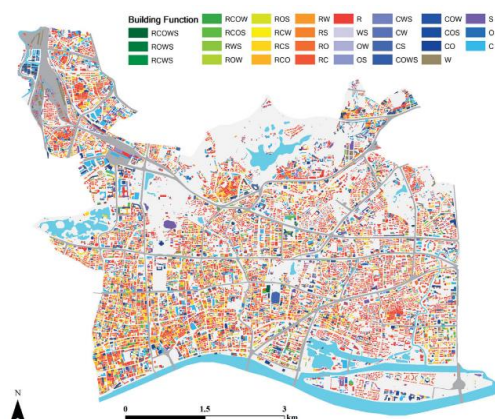
a. Bina Fonksiyonlarının OSM Ve Sosyal Medya Platformları ile Desteklenmesi (Fonte vd., 2018)



b. OSM Platformu Örneği (Atwal vd., 2022).



c. Tencent EasyGo Platformu Örneği (Gao vd., 2021).



d. Gaode Map Platformu Örneği (Deng vd., 2022).

Şekil 8. Platform Örnekleri ve Bina Fonksiyon Tespiti.

Zheng vd. (2024) ise Çin'in 38 kentinde OSM, Amap, Baidu ve resmi verileri XGBoost algoritmasıyla bütünleştirerek 39 milyon binayı sınıflandırmıştır. Benzer biçimde, şehircilik çalışmalarında sıklıkla kullanılan OSM verileri, binaların geometrik özellikleri hakkında temel bilgiler sunmasına rağmen, kullanım kategorileri gibi açıklayıcı öznitelikler açısından genellikle yetersizdir. Bunu eksikliği gidermek için çalışmalar yapılmıştır (Memduhoglu & Basaraner, 2024). Son yıllarda emlak ilanları, sosyal medya içerikleri ve konum tabanlı uygulamalardan elde edilen metin verileri de bina fonksiyonlarının belirlenmesinde kullanılmakta; doğal dil işleme teknikleriyle bu verilerden konut, ticaret, sağlık ve eğitim gibi fonksiyonlar otomatik biçimde çıkarılabilmektedir (Zhang vd., 2025).

Dijital veri kaynakları, yalnızca mekânsal analizlerde değil; afet yönetimi, ulaşım planlaması ve toplumsal karar süreçlerinde de etkin biçimde kullanılmaktadır. Acil durum yönetiminde uygulama programlama arayüzleri, sistemler arası veri paylaşımını kolaylaştırarak önemli bir rol üstlenmektedir. Örneğin, Google Maps API trafik yoğunluğu, yol kapanmaları ve alternatif güzergâhlar hakkında gerçek zamanlı bilgiler sunarak afet yönetimi uygulamalarında yaygın biçimde kullanılmaktadır. Salazar-Carrillo vd. (2021), Mexico City'de Twitter verilerini Destek Vektör Makinesi (DVM) algoritmasıyla analiz etmiş ve sonuçları web tabanlı CBS ortamında görselleştirerek trafik olaylarını haritalandırmıştır. Çalışma, sosyal medya verilerinin ulaşım izleme ve yönetiminde düşük maliyetli ve etkili bir araç olduğunu göstermektedir.

Web kazıma ve sosyal medya verilerinin birlikte değerlendirilmesi, kentlerdeki yapısal dönüşüm, kullanıcı davranışları ve sosyal etkileşimlerin anlaşılmasında yenilikçi bir yaklaşım sunmaktadır. Ensari ve Kobaş (2018), web kazıma yöntemlerini kentsel tasarım süreçlerinde kullanıcı geri bildirimlerinin değerlendirilmesinde kullanarak kanıta dayalı politika geliştirme potansiyelini vurgulamıştır. Liu vd. (2018), Çin merkezli dianping.com verilerini kullanarak metro istasyonlarının etkinliğini analiz etmiş; Nishimura vd. (2017) ise kullanıcı güzergâhlarını güvenlik ve gürültü gibi çevresel faktörlerle ilişkilendirerek kent içi davranış örüntülerini belirlemiştir.

Afet ve kriz dönemlerinde dijital veriler, toplumsal tepkilerin ve davranışların anlaşılmasında değerli bilgiler sağlamaktadır. Zha vd. (2023), Covid-19 sürecinde Wuhan'daki ulaşım politikalarına yönelik halk tepkilerini Weibo verilerini kullanarak duygu analizi ile Jaccard mesafesi yöntemleri ni kullanarak politika etkilerini değerlendirmiştir. Dijital veriler aynı zamanda katılımcı planlamayı desteklemekte, vatandaşların karar süreçlerine dâhil olmasını kolaylaştırmaktadır. Lin ve Kant (2021), Hollanda'da bisiklet altyapısına ilişkin halk görüşlerini Facebook aracılığıyla toplamış; içerik analizi ve web kazıma teknikleriyle değerlendirmiştir. Benzer biçimde, Chen vd. (2022) Çin'deki atık ayrıştırma politikasının yerel düzeydeki uygulamalarını sosyal medya verileriyle incelemiş ve merkezî politikaların toplumsal yansımalarını göstermiştir.

Web tabanlı dijital veri kaynakları, günümüz şehircilik çalışmalarında mekânsal analizlerin doğruluğunu ve güncelliğini destekleyen bileşenler haline gelmiştir. CBS, veri madenciliği, makine öğrenmesi ve doğal dil işleme yöntemleriyle bütünleştirilen bu veriler; bina fonksiyonlarının belirlenmesi, afet yönetimi, ulaşım planlaması ve katılımcı şehircilik gibi alanlarda önemli katkılar sağlamaktadır. Farklı veri türlerinin entegrasyonu, kentsel dinamiklerin daha bütüncül biçimde değerlendirilmesini mümkün kılmakta ve karar destek sistemlerinin bilimsel temelli biçimde güçlendirilmesine olanak tanımaktadır.

SONUÇLAR VE ÖNERİLER

Yapılan çalışmada, web tabanlı dijital veri kaynaklarının çeşitliliği, elde edilme yöntemleri ve bu verilerin şehircilik çalışmalarındaki kullanım alanları dikkate alınmıştır. Literatürde, web kazıma teknikleri ve uygulama programlama arayüzü tabanlı veri toplama yöntemleri gibi dijital veri elde etme yaklaşımlarının; ayrıca veri madenciliği gibi analiz yöntemlerinin, özellikle kentsel planlama, afet yönetimi, çevresel izleme ve ulaşım analizi gibi alanlarda önemli fırsatlar sunduğu tespit edilmiştir. Yapılan çalışmalar, web tabanlı dijital verilerin güncellik ve erişilebilirlik açısından önemli avantajlar sağladığını vurgulamakla birlikte; veri bütünlüğü, standardizasyon eksiklikleri ve gizlilik kaygıları gibi bazı sınırlamaların da altını çizmektedir. Ayrıca verilerin çok farklı kaynaklar tarafından üretilmesi veri entegrasyonu süreçlerinde ek zorluklar yaratabilmektedir.

Web tabanlı dijital veriler şehircilik çalışmalarında mekânsal planlama, ulaşım analizi, afet yönetimi, çevresel izleme, kamusal alanların değerlendirilmesi, ekonomik faaliyetlerin izlenmesi ve toplumsal eğilimlerin analizi gibi birçok uygulama alanında kullanılmaktadır. Özellikle kentsel büyüme, arazi kullanımı, bina fonksiyonlarının belirlenmesi ve nüfus hareketliliği gibi konularda web tabanlı dijital veri kaynaklarından elde edilen verilerin hem karar vericiler hem de araştırmacılar açısından stratejik öneme sahip olduğu görülmektedir. Ancak, bu verilerin doğruluğu, mekânsal çözünürlüğü gibi niteliksel özelliklerinin, şehir planlama süreçlerine etkin biçimde entegre edilebilmesi için dikkatle değerlendirilmesi gerekmektedir. Mevcut araştırmalar, dijital verilerin sağladığı hızlı ve geniş kapsamlı bilgiye rağmen, kimi durumlarda eksik, hatalı ya da önyargılı olabileceğini ve bu durumun analiz sonuçlarının güvenilirliğini olumsuz etkileyebileceğini göstermektedir. Bu nedenle, dijital veri kaynaklarının geleneksel veri setleriyle birlikte kullanılması, çoklu kaynaklar arasında çapraz doğrulama yapılması ve mekânsal veri doğruluğunu artırmaya yönelik yöntemlerin geliştirilmesi önemli bir gereklilik olarak öne çıkmaktadır.

Konu hakkında yapılacak çalışmalarda, özellikle dijital veri kaynaklarının kalite değerlendirmesi, veri bütünlüğü analizi ve etik veri kullanımı konularında derinleştirilmesi büyük önem arz etmektedir. Veri standartlarının geliştirilmesi, veri paylaşım protokollerinin oluşturulması ve açık veri kullanımının desteklenmesi, çalışmalarda web tabanlı dijital veri kullanımının etkinliğini artıracak temel adımlar arasında yer almaktadır. Ayrıca, makine öğrenmesi, yapay zekâ ve ileri veri çözümleme yöntemleri gibi yeni nesil teknolojilerin dijital veri kaynaklarıyla entegrasyonu ile hızlı ve doğru veri işleme süreçleri desteklenebileceği gibi daha öngörülebilir ve esnek şehir yönetim modelleri de geliştirilebilecektir. Akıllı şehir uygulamaları kapsamında, gerçek zamanlı veri izleme, mekânsal-zamansal analizlerin güçlendirilmesi ve vatandaş odaklı veri kullanım senaryolarının çoğaltılması, dijital veri kaynaklarının sunduğu potansiyelin daha verimli biçimde kullanılmasını sağlayacaktır.

Sonuç olarak, dijital veri kaynaklarının şehir alanlarında kullanımı, şehirlerin daha yaşanabilir, dirençli ve sürdürülebilir hâle getirilmesi açısından stratejik bir öneme sahiptir. Ancak bu potansiyelin tam anlamıyla gerçekleştirilebilmesi için, veri kalitesi, erişim, etik ve entegrasyon konularında sistematik çözümler geliştirilmesi gerekmektedir. Dijital veri kaynaklarının sunduğu fırsatların yanı sıra barındırdığı risklerin de göz önünde bulundurulması hem akademik çalışmaların hem de pratik şehir yönetimi uygulamalarının başarısını doğrudan etkileyen bir faktör olacaktır.

Etik Standartlara Uyum

Çıkar Çatışması

Yazarlar, bu çalışmada herhangi bir çıkar çatışmasının bulunmadığını beyan eder.

Etik Kurul İzni

Bu çalışma için etik kurul iznine gerek yoktur. Buna ilişkin ıslak imzalı onam formu, makale süreç dosyasına eklenmiştir.

Finansal Destek

Finansal destek bulunmamaktadır.

Teşekkür

Teşekkür edilecek kişi veya kurum bulunmamaktadır.

KAYNAKÇA:

- Atwal, K. S., Anderson, T., Pfoser, D., & Züfle, A. (2022). Predicting building types using OpenStreetMap. *Scientific Reports*, 12(1), 19976.
- Brindley, P., Goulding, J., & Wilson, M. L. (2018). Generating vague neighbourhoods through data mining of passive web data. *International Journal of Geographical Information Science*, 32(3), 498–523.
- Burrough, P. A., & McDonnell, R. A. (1998). *Principles of geographical information systems*. Oxford University Press Inc., New York.
- Calafiore, A., Palmer, G., Comber, S., Arribas-Bel, D., & Singleton, A. (2021). A geographic data science framework for the functional and contextual analysis of human dynamics within global cities. *Computers, Environment and Urban Systems*, 85, 101539.
- Chen, F., Hou, J., Gu, X., Hou, J., Pan, Q., & Tang, Z. (2022). Research on temporal and spatial evolution of public's response to the mandatory waste separation policy based on big data mining. *Sustainable Production and Consumption*, 31, 236–248.

- Chow, T. E., Schuermann, R. T., Ngu, A. H., & Dahal, K. R. (2018). Spatial mining of migration patterns from web demographics. *International Journal of Geographical Information Science*, 32(10), 1977–1998. <https://doi.org/10.1080/13658816.2018.1470633>.
- Deng, Y., Chen, R., Yang, J., Li, Y., Jiang, H., Liao, W., & Sun, M. (2022). Identify urban building functions with multisource data: A case study in Guangzhou, China. *International Journal of Geographical Information Science*, 36(10), 2060–2085.
- Dolu, D. G., & Şen, A. (2024). Coğrafi referanslı sosyal medya verilerinden doğal dil işleme ve derin öğrenme kullanılarak duygu analizi: Kahramanmaraş depremleri örneği. *Journal of Geodesy and Geoinformation*, 11(1), 51–67. <https://doi.org/10.9733/JGG.2024R0005.T>.
- Ensari, E., & Kobaş, B. (2018). Web scraping and mapping urban data to support urban design decisions. *A/Z: ITU Journal of Faculty of Architecture*, 15(1), 5–21. <https://doi.org/10.5505/itujfa.2018.40360>.
- Fonte, C. C., Minghini, M., Antoniou, V., Patriarca, J., & See, L. (2018). Classification of building function using available sources of VGI. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 209–215.
- Gao, F., Huang, G., Li, S., Huang, Z., & Chai, L. (2021). Integrating the eigendecomposition approach and k-means clustering for inferring building functions with location-based social media data. *ISPRS International Journal of Geo-Information*, 10(12), 834.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>.
- Haan, K. (2025, 24 Ekim). *Top Website Statistics For 2025*. Forbes. <https://www.forbes.com/advisor/business/software/website-statistics/>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann Publishers is an imprint of Elsevier.
- Jo, T. (2024). *Text Mining: Concepts, implementation, and big data challenge*. (2nd ed.) Springer.
- Kilic, B., & Gülgen, F. (2020). Investigating the quality of reverse geocoding services using text similarity techniques and logistic regression analysis. *Cartography and Geographic Information Science*, 47(4), 336–349. <https://doi.org/10.1080/15230406.2020.1746198>.
- Küçük, K. C., & Anbaroğlu, B. (2019). OpenStreetMap binalarının mekânsal doğruluğunun analizi. *Türkiye Coğrafi Bilgi Sistemleri Dergisi*, 1(1), 5–13.
- Kuşak, L. (2010). *Web ortamında mekânsal veri kazanımı (spatial data gain from web)* (Yayın No. 350628) [Doktora Tezi, İstanbul Teknik Üniversitesi]. YÖK Ulusal Tez Merkezi.
- Laender, A. H. F., Ribeiro-Neto, B. A., Da Silva, A. S., & Teixeira, J. S. (2002). A brief survey of web data extraction tools. *ACM SIGMOD Record*, 31(2), 84–93.
- Li, D., Wang, S., & Li, D. (2015). *Spatial data mining: Theory and application* (1st ed.) Springer
- Lin, Y., & Kant, S. (2021). Using social media for citizen participation: Contexts, empowerment, and inclusion. *Sustainability*, 13(12), 6635. <https://doi.org/10.3390/su13126635>.
- Liu, X., Macedo, J., Zhou, T., Shen, L., Liao, Y., & Zhou, Y. (2018). Evaluation of the utility efficiency of subway stations based on spatial information from public social media. *Habitat International*, 79, 10–17.

- Makhabel, B., Mishra, P., Danneman, N., & Heimann, R. (2017). *R: Mining spatial, text, web, and social media data* (1st ed). Packt Publishing.
- Makice, K. (2009). *Twitter API Up and Running*. O'Reilly Media.
- Martins, B. E. da G. M. (2008). *Geographically aware web text mining* [Doctoral dissertation, Universidade de Lisboa].
- Memduhoğlu, A. (2024). 6 Şubat depremlerinden etkilenen şehirlere ait OpenStreetMap verilerinin niceliksel ve semantik analizi. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 13(4), 1264–1276. <https://doi.org/10.28948/ngmuh.1476998>.
- Memduhoglu, A., & Basaraner, M. (2024). Semantic enrichment of building functions through geospatial data integration and ontological inference. *Environment and Planning B: Urban Analytics and City Science*, 51(4), 923–938. <https://doi.org/10.1177/23998083231206165>.
- Mitchell, R. (2018). *Web scraping with python collecting more data from the modern web* (2nd ed.). O'Reilly Media.
- Molina-Campoverde, J. J., Rivera-Campoverde, N., Molina Campoverde, P. A., & Bermeo Naula, A. K. (2024). Urban mobility pattern detection: development of a classification algorithm based on machine learning and GPS. *Sensors*, 24(12), 3884.
- Mueller, J. P. (2004). *Mining amazon web services: Building applications with the Amazon API*. Sybex.
- Nishimura, T., Nishida, K., Toda, H., & Sawada, H. (2017). Social media knows what road it is: Quantifying Road characteristics with geo-tagged posts. *Social Network Analysis and Mining*, 7(1), 57. <https://doi.org/10.1007/s13278-017-0473-y>.
- Özbek, T. N. N., & Ercoskun, O. (2023). Planlamada açık veri portalları/etkileşimli haritalar: İzmir (Türkiye) - Sidney (Avustralya) örneği. *Bilişim Teknolojileri Dergisi*, 16(3), 207–220. <https://doi.org/10.17671/gazibtd.1246353>.
- Petroutsos, E. (2014). *Google Maps: Power tools for maximizing the API*. McGraw Hill.
- Rajnish, D. R., & Srivastava, D. M. (2023). *Web data mining with Python: Discover and extract information from the web using Python*. BPB Publications.
- Richardson, L. (2025, 28 Nisan). *Beautiful Soup*. <https://www.crummy.com/software/BeautifulSoup>
- Salazar-Carrillo, J., Torres-Ruiz, M., Davis, C. A., Quintero, R., Moreno-Ibarra, M., & Guzmán, G. (2021). Traffic congestion analysis based on a Web-GIS and data mining of traffic events from Twitter. *Sensors*, 21(9), 2964.
- Scrapy. (2025, 28 Nisan). *Scrapy*. <https://scrapy.org>
- Serere, H. N., Resch, B., & Havas, C. R. (2023). Enhanced geocoding precision for location inference of tweet text using spaCy, Nominatim and Google Maps. A comparative analysis of the influence of data selection. *PLOS ONE*, 18(3), e0282942.
- Setiawan, E. B., Putera Husein, G., & Setiyadi, A. (2023). Travel route recommendation system based on weather prediction and geolocation technology. In *2023 International Conference on Informatics Engineering, Science & Technology (INCITEST)*, pp. 1–7. IEEE <https://doi.org/10.1109/INCITEST59455.2023.10395920>

- Taşkanat, T., Karaağaç, A., Beşdok, E., & Bostancı, B. (2018). Kentsel sorunların yönetimi için bir gönüllü coğrafi bilgi mobil uygulaması geliştirilmesi. *Geomatik*, 3(1), 84–91. <https://doi.org/10.29128/geomatik.371144>
- Tomlison, R. F. (1974). *The application of electronic computing methods and techniques to the storage, compilation, and assessment of mapped data* [Doctoral dissertation, University of London].
- Üsküplü, T., & Çolakoğlu, B. (2019). Sosyal ağ verileri ve mekân dizim analizlerinin kentsel stratejiler geliştirmede kullanımı: Kadıköy örneği. *Megaron*, 14(2), 269-278. <https://doi.org/10.14744/megaron.2019.58569>
- Vaggione, P. (2014). *Urban planning for city leaders* (2nd ed., Vol. 1). United Nations Human Settlements Programme (UN-Habitat). <https://unhabitat.org/urban-planning-for-city-leaders-0>
- Wang, Y., Wang, T., Tsou, M.-H., Li, H., Jiang, W., & Guo, F. (2016). Mapping dynamic urban land use patterns with crowdsourced geo-tagged social media (Sina-Weibo) and commercial points of interest collections in Beijing, China. *Sustainability*, 8(11), 1202. <https://doi.org/10.3390/su8111202>
- Zha, W., Ye, Q., Li, J., & Ozbay, K. (2023). A social media data-driven analysis for transport policy response to the COVID-19 pandemic outbreak in Wuhan, China. *Transportation Research Part A: Policy and Practice*, 172, 103669. <https://doi.org/10.1016/j.tra.2023.103669>
- Zhang, Y., Zhao, H., & Long, Y. (2025). CMAB: A Multi-Attribute Building Dataset of China. *Scientific Data*, 12(1). <https://doi.org/10.1038/s41597-025-04730-5>
- Zheng, Y., Zhang, X., Ou, J., & Liu, X. (2024). Identifying building function using multisource data: A case study of China's three major urban agglomerations. *Sustainable Cities and Society*, 108, 105498. <https://doi.org/10.1016/j.scs.2024.105498>

EXTENDED SUMMARY

Research Problem:

This article examines the potential of web-based spatial digital data sources for urban research by reviewing web scraping and API-based data acquisition processes and their relationship with traditional data collection methods. It also highlights key challenges related to data integrity, privacy, and standardisation, and presents selected urban planning and management applications.

Research Questions:

What are the digital data sources containing spatial information, and how are they classified in the literature?
What methods are used to extract spatial data from digital sources?
For what purposes and in what application areas are these digital data sources used in urban research?
What are the advantages and limitations of digital data sources compared to traditional spatial data sources?
What are the main strengths and weaknesses of the methods used to extract spatial information?
What techniques can be applied to obtain location-based information?
Is it possible to determine building functions using digital data sources, and how is this addressed in the literature?

Literature Review:

As technology has advanced, web-based digital data extraction has evolved from manual copying and basic HTML parsing to automated scraping frameworks such as BeautifulSoup and Scrapy. In parallel, APIs provided

by commercial and social media platforms, as well as municipal and institutional open data portals, have become widespread, enabling easier access to large volumes of structured digital data.

In parallel with these developments, spatial digital data sources have evolved rapidly, from early satellite missions such as Landsat-1 to the expansion of GIS and GPS technologies in the late twentieth century. The emergence of internet-based mapping services, followed by user-contributed platforms such as OpenStreetMap and Google Earth in the mid-2000s, created a hybrid data ecosystem in which spatial information is increasingly produced and shared beyond institutional structures.

Within this technological transformation, urban studies have increasingly utilized web-based spatial data. Sources such as OpenStreetMap (OSM), location-based services, Point of Interest (POI) platforms, user review sites, real estate and tourism portals, social media, and municipal open data are widely used to analyse land use and building functions, functional zones, mobility patterns, and environmental or disaster risks. Recent studies also demonstrate the integration of OSM with official records to improve semantic consistency, and the use of digital traces (e.g., search or social media data) to capture behavioural and mobility dynamics during events.

In addition, APIs such as Google Maps and OpenWeatherMap are used in studies to obtain real-time transportation and climate data. In recent years, machine learning and natural language processing techniques have been used extensively for tasks such as accessing digitized data, geocoding, POI integration, and function extraction.

Despite these developments, the literature highlights persistent methodological and ethical challenges in using web-based geographic data, including spatial inaccuracies, limited standardization, uncontrolled user-generated content, and semantic inconsistencies across platforms. In addition, privacy and personal data protection regulations such as the European Union's General Data Protection Regulation and Türkiye's Personal Data Protection Law (Law No. 6698), as well as licensing conditions, robots.txt restrictions, and usage permissions, may limit data access and reuse. Therefore, researchers emphasize the need for systematic quality assessment, compatibility checks with official datasets, and responsible use guidelines.

Methodology:

This study is a conceptual and literature-based research that examines the use of web-based spatial digital data in urban research. PhD and master's theses, scientific articles, books, and relevant web-based sources were critically reviewed, and the main arguments derived from these sources were synthesized. The study evaluates the theoretical background of web scraping, API-based data acquisition, and data mining approaches in the context of urban studies.

Findings:

This review presents evidence that web-based spatial digital data is becoming increasingly important in urban research. Overall, these data sources complement traditional datasets by offering high temporal resolution, rich semantic content, and broad spatial coverage from neighborhood to global scales. Social media data, sensor-based APIs, and the real-time data features provided by platform services enable the near real-time monitoring of urban processes such as traffic density, disaster behavior, and user mobility. Furthermore, POI data, user reviews, business descriptions, and text content provide rich semantic contributions to applications such as building function extraction, usage intensity estimation, and functional zoning. The scalability of web-based data is also noteworthy. This enables multi-scale analyses ranging from the neighborhood level to national or international comparisons.

Discussion:

Studies also show that web-based spatial digital data sources have significant structural limitations despite their strengths. Issues of data integrity and representation are frequently highlighted on user-contributed platforms (such as OSM, social media, and business websites). The overrepresentation of some areas and the near invisibility of others, coupled with varying timeliness across platforms, creates significant uncertainty when these data sources are used alone in spatial analyses.

The studies also show that the lack of standardization in the digital data ecosystem and cross-platform incompatibilities are widespread problems. These limitations indicate that single-source analyses are more susceptible to structural errors and that hybrid approaches that combine multiple data sources are increasingly necessary. Therefore, the efficient use of web-based spatial digital data requires a comprehensive approach that addresses data integration, standardization, ethical compliance, and preprocessing processes together.

Conclusion:

This study comprehensively evaluates the diversity of web-based digital data sources with spatial content, the methods used to obtain these data, and their potential applications in urban planning. The studies examined demonstrate that web-based spatial data make significant contributions to numerous urban applications, including spatial planning, disaster management, environmental monitoring, and transportation analysis.

Future studies in this area should focus on developing standards for the quality assessment of digital data, enhancing data integrity, clarifying ethical and institutional frameworks, and strengthening mechanisms that support open data sharing. The combination of machine learning, artificial intelligence, and real-time spatial analytics with web-based spatial data sources can make significant contributions to the creation of more sustainable and citizen-centered urban management models.