

İSTATİSTİK HESAPLAMALARI LABORATUVARI

Araştırma hedeflerinden veri kaynakları tarama ve hazırlık çalışması, araştırmacılar ve işletmeler için faydalı bazı veri kaynaklarını göstermek amacı ile oluşturulmuştur.

Veri kaynakları tarama ve hazırlık çalışma ekibi:

Dr.Öğr.Üyesi Mükerrerrem Bahar Başkır

Araş.Gör. Ayşenur Akın Vargeloğlu

YASAL UYARI: Burada verilen dış bağlantıların içerisinde yer alan bilgilerin ve veri setlerinin güncelliğinden ve doğruluğundan ilgili siteler sorumludur. Bu hususta, veri kaynağı taraması ve hazırlığını gerçekleştiren kişiler ve bağlı oldukları kurumlar hiçbir şekilde sorumlu tutulamaz. Bu sitede yer alan veri kaynakları tarama ve hazırlık bilgileri ilgili çalışma ekibi tarafından yapay zeka, makine öğrenme, veri bilimi araştırmacıları ve işletmeler için genel bilgi elde edilebilmesi amacı ile hazırlanmıştır.

VERİ KAYNAKLARI

Araştırmacılar ve işletmeler için faydalı bazı veri kaynakları, genel bilgileri ve linkleri ile birlikte aşağıda listelenmektedir.

A. Multidisipliner alanlarda veri setleri içeren bazı kaynaklar:

Google Dataset Search: Bilim insanları ve araştırmacılar için 25 milyondan fazla veri setine en hızlı erişimi sağlayan veri depolarından biridir. Google Akademik servisinin bir uzantısı olarak bilimsel çalışmalarda kullanılan veri setlerine kolay erişime olanak sağlamaktadır.

Link: <https://datasetsearch.research.google.com/>

Kaggle: Bir çok alana özgü veri ihtiyacını karşılayan etkileşimli veri bilimi sitesidir. Veri bilimi ve makine öğrenmesine ilgi duyan araştırmacıları bir araya getiren platform, çeşitli konularda interaktif çalışma olanağı sağlamaktadır.

Link: <https://www.kaggle.com/>

UCI Machine Learning: En eski ve yaygın kullanılan veri depolarından biridir. Çeşitli kategorilerde ve veri tiplerinde (tek ve çok değişkenli zaman serisi, sınıflandırma, regresyon vb.) veri setleri kayıt olmadan doğrudan UCI Machine Learning veri havuzundan elde edilebilir.

Link: <https://archive.ics.uci.edu/ml/datasets.php>

World Bank Open Data: Eğitim, cinsiyet, sağlık, nüfus demografisi ve dünyanın dört bir yanından çok sayıda ekonomik verilere açık erişim sağlamaktadır.

Link: <https://databank.worldbank.org/home.aspx>

TÜİK: Türkiye’deki istatistiksel verilerin toplanıp paylaşılmasına imkan sağlayan önemli veri kaynaklarından biridir. Türk toplumunun bir çok alandaki (sosyal ve demografik, iş, tarım, çevre vs.) veri setlerine kolay ve hızlı erişimine olanak sağlamaktadır.

Link: <http://www.tuik.gov.tr/Start.do>

Veri Kaynağı: Türkiye’nin GSYH miktarı, eğitim, maden, vb. katagorilerde açık kaynaklı, anlamlı ve okunabilir verileri setlerine erişim sağlayan veri kaynağıdır.

Link: <https://www.verikaynagi.com/>

OECD Data: Araştırmacılara farklı kategorilerde (sağlık, tarım, ekonomi, finans, eğitim, sosyal vb.) veri setleri sunan en gelişmiş veri kaynaklarından biridir.

Link: <https://data.oecd.org/>

Microsoft Dataset: Matematik, sağlık, bilgisayar, sosyal, fizik vb. çeşitli katagorilerde veri setlerine açık erişime izin veren geniş kapsamlı veri kaynağıdır.

Link: <https://msropendata.com/>

EuroStat: AB’nin Lüksemburg’da bulunan istatistik ofisidir. İşletmeler, hükümetler, eğitim kurumları, tarım, balıkçılık, uluslararası ticaret vb. bir çok alanda veri sağlamaktadır. Araştırmacılar tarafından en sık kullanılan veri kaynaklarından biridir.

Link: <https://ec.europa.eu/eurostat/data/database>

Data.gov: ABD Hükümetinin açık verilerinin çoğunu içeren veri kaynağıdır. Araştırmalarınızda veya belirli bir kitleyi hedefleyen mobil uygulamalar oluşturmaya yardımcı olacak araçlar ve kaynaklar içermektedir. Kapsamlı filtreleme özellikleri ile iklim, tüketici, ekosistemler, eğitim, enerji, finans, sağlık, yerel yönetim, üretim, denizcilik gibi yaklaşık 15 farklı kategoride veri setleri içermekte ve araştırmacılara geniş bir veri kaynağı sunmaktadır.

Link: <https://www.data.gov/>

Awesome Public Datasets Collection: Tarım, biyoloji, ekonomi, eğitim, makine öğrenmesi gibi çeşitli alanlarda veriler içeren ve çoğuna ücretsiz erişim sağlayan GitHub sayfasıdır.

Link: <https://github.com/awesomedata/awesome-public-datasets>

DataStock: İnternet ortamında temiz ve kullanıma hazır veri setleri elde etmenizi sağlayan veri havuzudur. E-ticaret, seyahat vb. alanlarında veri setleri içermektedir.

Link: <https://datastock.shop/>

Nature: Dünyanın en iyi dergilerinden olan Nature, bilimsel ve değerli veri setleri ve açıklamaları için açık erişim olanağı sağlamaktadır.

Link: <https://www.nature.com/sdata/>

The UK Data Service: İngilterenin sosyal ekonomik ve demografik veri setlerini içerir.

Link: <https://ukdataservice.ac.uk/>

Amazon Web Services public datasets : Araştırmacıların AWS kaynakları aracılığıyla paylaşılan bir çok veri setine ulaşmasını sağlar. NASA'nın uydu görüntüleri ve 1000 Genom Projesi gibi bir çok kamunun erişimine açık veri setlerini içerir.

Link: <https://registry.opendata.aws/>

European Union Open Data Portal : Avrupa Birliği kurumları tarafından paylaşılan verilere erişim noktasıdır. Araştırmacılar, 20 farklı kategoride veri setlerine kolay ve hızlı erişim sağlayabilir.

Link: <https://data.europa.eu/euodp/en/data/>

The US National Center for Education Statistics: (ABD Eğitim İstatistikleri Merkezi) eğitim ve ilgili alanlarla ilgili büyük bir veri havuzuna sahiptir.

Link: <https://nces.ed.gov/>

New York Times : New York Time gazetesi verilerini kullanıma açtığı sitedir.

Link: <https://www.nytimes.com/column/raw-data>

National Climatic Data Center: İklimsel bir çok veri setini içeren sitedir.

Link: <https://www.ncdc.noaa.gov/>

The CIA World Factbook : CIA tarafından elde edilen ve paylaşımına açık 267 ülkenin tarih, nüfus ve ekonomi gibi veri setlerini içermektedir.

Link: <https://www.cia.gov/library/publications/the-world-factbook/>

CMU Datasets: Yüksek kalite verileri içermesi açısından dünyanın en iyi veri kaynaklarından biridir.

Link: <http://mocap.cs.cmu.edu/>

B. Görsel veri - sosyal medya veri setleri içeren bazı kaynaklar:

Data USA: Paylaşılan ABD verilerini toplamasına, analiz etmesine, görselleştirmesine ve veri paylaşımına olanak sağlayan ücretsiz bir platformdur

Link: <https://datausa.io/>

Açık Veri ve Veri Gazeteciliği : Veri gazeteciliği, açık veri, veri görselleştirme alanlarında çalışmalarını yürüten ve araştırmacılara çeşitli konularda veri setleri sunan veri kaynağıdır.

Link: <http://www.verigazeteciligi.com/>

Visual data : Görsel veri setleri içermektedir.

Link: <https://www.visualdatamedia.com/>

Google Books Ngrams: Google'ın 1500 ve 2008 yılları arası basılmış kaynaklarda bulunan sözcüklerin n-gram sayılarını kullanarak taratılmasını sağlayan ve bu sözcüklerin hangi yıl kaç defa kullanıldığını gösteren çevrimiçi arama motorudur.

Link: <https://books.google.com/ngrams>

Google Trends: Google aramalarında kelime ve cümle sorgularının hangi ülke ve şehirlerde ne sıklıkla taratıldığının istatistiksel bilgisi ve yorumlamalarına olanak sağlayan servistir.

Link: <https://trends.google.com.tr/trends/?geo=TR>

Topsy : Facebook ve Twitter gibi sosyal medya odaklı oldukça büyük bir veri setine açık erişim sağlamaktadır.

Link: <http://topsy.thisisthebrigade.com/>

Facebook Graph: Gizlilik ayarı kapalı olmayan facebook kullanıcıları ile ilgili çeşitli veri setleri içermektedir.

Link: <https://developers.facebook.com/docs/graph-api/>

C. Ekonomi alanında veri setleri içeren bazı kaynaklar:

IMF Data: Uluslararası para fonu, borç oranları, döviz rezervleri, emtia fiyatları ve yatırımlar hakkındaki veri setleri erişimine olanak sağlamaktadır.

Link: <https://www.imf.org/en/Data>

Financial Times Market Data: Hisse senedi, fiyat endeksleri, emtialar ve döviz de dahil olmak üzere dünyanın dört bir yanından finansal piyasalar hakkında güncel veri setleri içermektedir.

Link: <https://markets.ft.com/data>

American Economic Association (AEA): ABD makroekonomik verilerini bulmak için kullanılan veri kaynaklarından biridir.

Link: <https://www.aeaweb.org/>

Quandl: Ekonomik, finansal ve alternatif veri setleri için kullanılan popüler veri kaynağıdır.

Link: <https://www.quandl.com/>

US Census Bureau : ABD vatandaşlarının ekonomisi hakkında veri üretiminden sorumlu olmakla birlikte, sağlık konusunda en güncel verileri içeren site, veri setlerini kullanarak elde edilen görselleştirmelerde erişim sağlamaktadır.

Link: <https://www.census.gov/>

Google Finance: Son 40 yıla dair çeşitli ülkelerin ekonomik verilerini içeren sitedir.

Link: <https://www.google.com/finance>

D. Sağlık ve ilgili alanlarda veri setleri içeren bazı kaynaklar:

Healthdata.gov : ABD halkının paylaşılan çeşitli sağlık verilerine (çevre sağlığı, tıbbi cihazlar, sosyal hizmetler, toplum sağlığı, zihinsel sağlık ve madde bağımlılığı vb.) erişim sağlayan sitedir.

Link: <https://healthdata.gov/>

NHS Health and Social Care Information Centre : Sağlık ve ilgili konularda veri setlerine erişim kaynaklarından biridir.

Link: <https://www.gov.uk/government/organisations/health-and-social-care-information-centre>

Chronic disease data: ABD'deki kronik hastalıklarla ilgili veri setleri içerir.

Link: <https://www.cdc.gov/chronicdisease/data/index.htm>

The National Center for Biotechnology Information: Sağlık, genomik ve biyoteknoloji vb alanlarda her türlü geniş bilginin yanı sıra veritabanları içeren önemli bir kaynaktır.

Link: <https://www.ncbi.nlm.nih.gov/>

E. Bazı açık erişimli sağlık veri kaynakları web adresleri:

- <https://nihcc.app.box.com/v/ChestXray-NIHCC>
- <https://www.cancerimagingarchive.net/>
- <https://medpix.nlm.nih.gov/home>
- <http://www.oasis-brains.org/>
- <http://adni.loni.usc.edu/>
- <https://fitbir.nih.gov/>
- <http://cecas.clemson.edu/~ahoover/stare/>
- <https://www.insight-journal.org/midas/community/view/21>

F. Uluslararası bazı devletlerle ilgili halka açık olarak paylaşılan veri kaynakları:

- [EU Open Data](#)
- [US Gov Data](#)
- [New Zealand's Government Dataset.](#)
- [Indian Government Dataset.](#)
- <https://www.opendatani.gov.uk/>