



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



VERİ BİLİMİ VE ANALİTİĐİ EL KİTAPÇIĐI

Veriden Deđere Yolculuk ve Stratejik Uygulamalar

Ankara • 2025



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



ÖNSÖZ

Bu el kitapçığı, KızBaşına Genç Yetenek Geliştirme Programı kapsamında veri bilimi alanında eğitim alan katılımcılar için hazırlanmıştır. İçeriđi, bu alanda kariyer planlayan ya da veri ile çalışan herkesin başvurabileceđi biçimde tasarlanmıştır.

Veri bilimi, günümüzde sadece teknik bir disiplin olmaktan çıkıp, karar alma süreçlerinin merkezinde yer alan stratejik bir yetkinlik alanı haline gelmiştir. Bir belediyenin hizmet kalitesini iyileştirmesi, bir sivil toplum kuruluşunun toplumsal etkisini ölçmesi, bir şirketin müşteri ihtiyaçlarını anlaması ya da bir araştırmacının politika önerisi geliştirmesi; bunların tümü artık veriyle başlayan ve veriyle olgunlaşan süreçlerdir.

KızBaşına olarak bu rehberi hazırlarken iki temel inançtan hareket ettik: Birincisi, veri bilimi yetkinliklerinin doğru kullanıldığında toplumsal bir dönüşüm gücü taşıdığı. İkincisi, genç kadınların bu alanda yalnızca yararlanıcı değil, baş aktör olduđu. Impactathon etkinliđi boyunca program mezunlarının gerçek belediye verileriyle ürettikleri çözümler ve bu çözümlerin yerel yönetim süreçlerine yaptıđı somut katkılar, bu inancın yaşayan kanıtlarıdır.

Kitapçık; temel kavramlardan veri analitiđi yaşam döngüsüne, modern araçlardan büyük veri mimarilerine, etik yaklaşımlardan kariyer fırsatlarına uzanan kapsamlı bir çerçeve sunmaktadır. Her bölüm, teorik bilgiyi pratik uygulama senaryolarıyla birleştirecek biçimde kurgulanmıştır.

Bu kitapçığın, onu eline alan herkes için işlevsel bir başvuru kaynađına dönüşmesini umuyoruz.

KızBaşına Ekibi

Genç Yetenek Geliştirme Programı, KızBaşına tarafından Ankara Büyükşehir Belediyesi ortaklığıyla yürütölmekte olup Sivil Katılım Projesi kapsamında Avrupa Birliđi tarafından finanse edilmektedir. Proje, Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından Türkiye Belediyeler Birliđi ortaklığında, Dışişleri Bakanlığı Avrupa Birliđi Başkanlığı işbirliğinde yürütölen Sivil Katılım Hibe Programı kapsamında hayata geçirilmektedir. Bu yayın Avrupa Birliđi'nin mali desteđiyle Sivil Katılım Hibe Programı kapsamında hazırlanmıştır. Proje kapsamındaki tüm içerikler yalnızca KızBaşına sorumluluđundadır; UNDP ve Avrupa Birliđi'nin görüşlerini yansıtmamaktadır.



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



İÇİNDEKİLER

BÖLÜM 1: VERİ BİLİMİNE GİRİŞ

- 1.1 Veri Çađı ve Stratejik Bilgiye Ulaşmak
- 1.2 Genç Kadınların Veri Bilimindeki Rolü

BÖLÜM 2: VERİ BİLİMİ NEDİR?

- 2.1 Disiplinlerarası Bir Yaklaşım
- 2.2 Veri Biliminin Temel Bileşenleri

BÖLÜM 3: VERİ ANALİTİĐİ SÜRECİ

- 3.1 CRISP-DM Metodolojisi
- 3.2 Veri Analitiđi Yaşam Döngüsü
- 3.3 MLOps Entegrasyonu

BÖLÜM 4: TEMEL VERİ ANALİZİ TEKNİKLERİ

- 4.1 İstatistiksel Analiz ve Hipotez Testleri
- 4.2 Makine Öğrenmesi Teknikleri
- 4.3 Veri Görselleştirme ve Hikaye Anlatıcılığı
- 4.4 Veritabanı Yönetimi: SQL ve NoSQL

BÖLÜM 5: BÜYÜK VERİ MİMARİLERİ

- 5.1 Büyük Veri Ekosistemi ve Bulut Platformları
- 5.2 Veri Entegrasyonu: ETL ve ELT

BÖLÜM 6: PROJE YÖNETİMİ VE ETİK

- 6.1 MLOps ile Süreç Yönetimi
- 6.2 Veri Etiđi ve Sorumlu Yapay Zeka

BÖLÜM 7: KARİYER YOLLARI VE GELECEK

- 7.1 Veri Bilimi Kariyer Olanakları
- 7.2 Geleceğin Trendleri

SONUÇ: VERİ ODAKLI LİDERLİK VE TOPLUMSAL ETKİ

REFERANSLAR



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



BÖLÜM 1: VERİ BİLİMİNE GİRİŞ

1.1 Veri Çağı ve Stratejik Bilgiye Ulaşmak

Günümüz dünyası, her geçen gün artan devasa veri yığınlarıyla karakterize edilmektedir. Bu veri yığınları, doğru analiz edildiğinde karar alma süreçlerini dönüştürebilecek, yeni iş modelleri yaratabilecek ve toplumsal sorunlara yenilikçi çözümler sunabilecek paha biçilmez içgörüler barındırır. Veri bilimi ve analitiği, bu ham veriyi anlamlı bilgiye ve nihayetinde stratejik değere dönüştürme disiplini.

Bir belediyenin hizmet planlamasını veriyle yönlendirmesi, bir sivil toplum kuruluşunun savunuculuk faaliyetinin etkisini ölçmesi, bir şirketin müşteri davranışını öngörmesi ya da bir araştırmacının kanıt temelli politika önerisi üretmesi; bunların tümü, ham veriyi anlamlı bir hikayeye dönüştürme becerisinden geçer. Veri bilimi, bu dönüşümün hem yöntemini hem de dilini sunar.

1.2 Genç Kadınların Veri Bilimindeki Rolü

Genç kadınların bu alandaki yetkinliklerini geliştirmesi, hem bireysel kariyerlerinde rekabet avantajı sağlamak hem de veri odaklı bir gelecek inşa etme potansiyeli sunarak toplumsal ve ekonomik kalkınmada öncü roller üstlenmelerini sağlamaktadır. Kapsayıcı bir veri ekosistemi, ancak farklı deneyim ve perspektiflerden gelen kişilerin sürece dahil olmasıyla mümkündür. Aksi halde algoritmaların önyargılı, veri setlerinin eksik temsilli ve üretilen çözümlerin tek boyutlu kalma riski yüksektir.

KızBaşına Genç Yetenek Geliştirme Programı, genç kadınları veri bilimi alanında donatarak onların sadece teknoloji tüketicisi değil, aynı zamanda üreticisi ve yönlendiricisi olmalarını hedeflemektedir. Bu kitapçık, programın eğitim içeriğini destekleyen ve veri bilimi yolculuğuna sağlam bir temel oluşturmayı amaçlayan kapsamlı bir kaynaktır.

Neden Veri Bilimi?

Veri bilimi, karar alma süreçlerini sezgiden kanıta taşıyan stratejik bir yetkinlik alanıdır. Kapsayıcı veri ekosistemi, farklı deneyimlerden gelen kişilerin sürece dahil olmasıyla mümkündür. Genç kadınların veri bilimine dahil olması, daha adil ve dengeli algoritmaların ön koşuludur. Sivil alandan kamuya, özel sektörden akademiye uzanan geniş bir kariyer yelpazesi sunar. Veriden anlam üretmek, hem bireysel kariyer hem de toplumsal etki için kritik bir araçtır.



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



BÖLÜM 2: VERİ BİLİMİ NEDİR?

2.1 Disiplinlerarası Bir Yaklaşım

Veri bilimi, verilerden bilgi ve içgörü elde etmek için bilimsel yöntemleri, süreçleri, algoritmaları ve sistemleri kullanan disiplinlerarası bir alandır. İstatistik, bilgisayar bilimi, makine öğrenmesi ve alan uzmanlığının kesişim noktasında yer alır. Modern veri bilimciler, karmaşık ve heterojen veri setlerini analiz ederek gelecekteki eğilimleri tahmin eder, karar alıcılara yol gösterir ve veri odaklı ürünler ile hizmetler geliştirir.

Bu disiplinlerarası yapı, veri bilimini sadece teknik bir alan olmaktan çıkarıp stratejik bir karar destek mekanizması haline getirmektedir. İyi bir veri bilimci, sadece kod yazmayı bilmek ya da istatistik bilmekle yetinmez; aynı zamanda iş alanını anlar, doğru soruları sorar ve elde ettiği bulguları farklı kitlelere etkili biçimde aktarır.

2.2 Veri Biliminin Temel Bileşenleri

Veri bilimi, farklı disiplinlerden gelen bilgi ve becerilerin bir araya gelmesiyle oluşur. Bir veri bilimcinin başarılı olabilmesi için aşağıdaki bileşenlerin her birinde belirli bir yetkinlik düzeyine sahip olması beklenir.

Bileşen	Açıklama	Temel Yetkinlikler
Matematik ve İstatistik	Veri analizi için temel teorik çerçeveyi sağlar (olasılık, hipotez testi, regresyon, çok değişkenli analiz).	İstatistiksel modelleme, hipotez testi, lineer cebir, kalkülüs.
Programlama	Veri işleme, depolama ve algoritma geliştirme için programlama becerileri (Python, R, SQL).	Python/R, SQL, veri yapıları, algoritmalar.
Makine Öğrenmesi	Verilerden öğrenen modeller geliştirme; tahmin ve sınıflandırma görevleri için algoritmalar.	MÖ algoritmaları, derin öğrenme, model değerlendirme.
Alan Uzmanlığı	Analiz edilen verinin bağlamını ve iş alanının dinamiklerini anlama, doğru soruları sorma.	İş analizi, problem çerçeveleme, sektörel bilgi.
İletişim ve Görselleştirme	İçgörülerini anlaşılır, etkili ve ikna edici biçimde sunma; veri hikaye anlatıcılığı.	Tableau, Power BI, Matplotlib, raporlama becerileri.



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



BÖLÜM 3: VERİ ANALİTİĞİ SÜRECİ

3.1 CRISP-DM Metodolojisi

Veri analitiği süreci, genellikle tekrarlayan ve birbirini besleyen adımlardan oluşur. Bu adımlar, ham verinin toplanmasından anlamlı içgörülerin elde edilmesine kadar uzanır. Sektörde yaygın olarak kullanılan **CRISP-DM (Cross-Industry Standard Process for Data Mining)** metodolojisi, bu süreci sistematik biçimde yönetmek için bir çerçeve sunar.

Günümüzde bu süreç, MLOps prensipleriyle entegre edilerek daha çevik, ölçeklenebilir ve sürdürülebilir hale getirilmektedir. Bu entegrasyon, modellerin geliştirme ortamından üretim ortamına sorunsuz bir geçiş yapmasını sağlar.

3.2 Veri Analitiği Yaşam Döngüsü

İş Anlayışı (Business Understanding)

Bu aşama, veri biliminin en kritik başlangıç noktasıdır. Hangi iş problemine çözüm aranıyor? Hangi hedeflere ulaşılmak isteniyor? Paydaşlarla yakın işbirliği içinde projenin kapsamı, başarı kriterleri, kısıtlamaları ve beklenen iş değeri netleştirilir. Doğru soruları sormak, projenin başarısı için temeldir. Örneğin, "Müşteri kaybını nasıl azaltabiliriz?" ya da "Hangi kampanya en etkili olacak?" gibi sorular bu aşamada belirlenir.

Veri Anlayışı (Data Understanding)

İlgili verilerin çeşitli kaynaklardan (veritabanları, API'ler, veri gölleri, web kazıma, sensörler) toplanması ve keşifsel veri analizi (EDA) ile verinin yapısının, kalitesinin, dağılımının ve potansiyel sorunlarının (eksik değerler, aykırı değerler, veri tipleri) anlaşılması bu aşamada gerçekleşir. Veri görselleştirmeleri ve istatistiksel özetler yoğun olarak kullanılır.

Veri Hazırlığı (Data Preparation)

Bu aşama, veri bilimcilerin zamanının büyük bir kısmını harcadığı yerdir. Eksik, hatalı veya tutarsız verilerin düzeltilmesi, aykırı değerlerin yönetimi, özellik mühendisliği (feature engineering), veri dönüştürme (normalizasyon, standardizasyon) ve formatlama işlemleri yapılır. Temiz ve iyi hazırlanmış veri, modelin başarısını doğrudan etkiler.

Modelleme (Modeling)

Hazırlanmış veri setleri üzerinde istatistiksel modeller, makine öğrenmesi algoritmaları veya derin öğrenme mimarileri kullanılarak verilerdeki ilişkiler ve eğilimler belirlenir. Farklı modellerin denenmesi, hiperparametre optimizasyonu ve en uygun olanın seçilmesi bu aşamada gerçekleşir.

Değerlendirme (Evaluation)



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



Modelin performansı iş hedefleri doğrultusunda değerlendirilir (dođruluk, kesinlik, geri çağırma, F1 skoru, ROC eğrisi, RMSE, MAE gibi metrikler kullanılarak). Modelin iş değeri, maliyet-fayda analizi ve etik etkileri de bu aşamada gözden geçirilir. Modelin genellenebilirliđi ve aşırı öğrenme (overfitting) durumu kontrol edilir.

Dađıtım (Deployment)

Başarılı bulunan modelin operasyonel ortama entegre edilmesi ve gerçek dünya verileri üzerinde çalışır hale getirilmesi bu aşamada yapılır. MLOps prensipleriyle (otomatikleştirilmiş dađıtım, versiyonlama, API entegrasyonları, konteynerizasyon) desteklenir.

İzleme ve Bakım (Monitoring & Maintenance)

Dađıtılan modelin performansının sürekli izlenmesi, veri kayması (data drift), model eskimesi (model decay) veya performans düşüşü gibi durumların tespiti ve modelin düzenli olarak güncellenmesi bu aşamanın temel işlevleridir. Geri bildirim döngüleri oluşturularak model sürekli iyileştirilir.



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



BÖLÜM 4: TEMEL VERİ ANALİZİ TEKNİKLERİ

Veri bilimcilerin kullandığı çok sayıda teknik ve araç bulunmaktadır. Bu bölümde en yaygın ve etkili olanlar detaylandırılacaktır.

4.1 İstatistiksel Analiz ve Hipotez Testleri

İstatistik, veri biliminin temelini oluşturur. Veri setlerini anlamak, ilişkileri keşfetmek ve çıkarımlar yapmak için istatistiksel yöntemler vazgeçilmezdir.

- **Tanımlayıcı İstatistikler:** Veri setinin temel özelliklerini özetleme. Ortalama, medyan, mod, standart sapma, varyans, çeyreklikler, frekans dağılımları, çarpıklık ve basıklık gibi ölçütler kullanılır.
- **Çıkarımsal İstatistikler:** Örneklem verilerinden yola çıkarak daha geniş bir popülasyon hakkında çıkarımlar yapma. Hipotez testi (t-testi, ANOVA, ki-kare), güven aralıkları ve korelasyon analizi bu kategoriye girer.
- **Regresyon Analizi:** Bağımlı ve bağımsız değişkenler arasındaki ilişkileri modelleme ve tahmin yapma. Lineer regresyon, lojistik regresyon, çoklu regresyon ve zaman serisi regresyonu gibi türleri vardır.

4.2 Makine Öğrenmesi Teknikleri

Veri biliminde makine öğrenmesi algoritmaları; tahmin, sınıflandırma, kümeleme ve boyut azaltma görevleri için yaygın olarak kullanılır. Detaylı bilgi için *Yapay Zeka ve Makine Öğrenmesi El Kitabı*'na bakılması önerilir.

- **Sınıflandırma Algoritmaları:** Veri noktalarını önceden tanımlanmış kategorilere ayırır. Karar Ağaçları, Random Forest, SVM, Lojistik Regresyon, Naive Bayes, KNN gibi algoritmalar kullanılır. Örneğin bir e-postanın spam olup olmadığını tahmin etmek.
- **Regresyon Algoritmaları:** Sürekli bir hedef değişkeni tahmin eder. Lineer Regresyon, Ridge, Lasso, Gradient Boosting Makineleri gibi algoritmalar bu kategoride yer alır. Örneğin bir evin fiyatını tahmin etmek.
- **Kümeleme Algoritmaları:** Etiketsiz veri setlerindeki benzer veri noktalarını gruplar. K-Means, Hiyerarşik Kümeleme, DBSCAN, Gaussian Mixture Models gibi algoritmalar kullanılır. Örneğin müşteri segmentasyonu.
- **Boyut Azaltma Teknikleri:** Veri setindeki değişken sayısını azaltarak gürültüyü giderir, görselleştirmeyi kolaylaştırır ve model performansını artırır. PCA, t-SNE, UMAP gibi teknikler kullanılır.

4.3 Veri Görselleştirme ve Hikaye Anlatıcılığı

Veri görselleştirme, karmaşık veri setlerindeki desenleri, eğilimleri ve aykırı değerleri anlaşılır, estetik ve etkili grafikler aracılığıyla sunma sanatıdır. Etkili görselleştirmeler, içgörülerin hızlı



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



kavranmasını sağlar ve karar alma süreçlerini destekler. Veri hikaye anlatıcılığı (data storytelling) ise görselleştirmeleri bir anlatı ile birleştirerek mesajın daha güçlü ve akılda kalıcı biçimde iletilmesini sağlar.

- **Grafik Türleri:** Bar grafikler, çizgi grafikler, dağılım grafikleri (scatter plots), ısı haritaları (heatmaps), kutu grafikleri (box plots), ağaç haritaları (treemaps), coğrafi haritalar ve ağ grafikleri. Her tür belirli bir veri ilişkisini vurgulamak için en uygundur.
- **Araçlar:** Python (Matplotlib, Seaborn, Plotly, Bokeh), R (ggplot2), Tableau, Power BI ve D3.js gibi güçlü görselleştirme kütüphaneleri kullanılır. Bu araçlar interaktif ve dinamik görselleştirmeler sunar.
- **Hikaye Anlatıcılığı Prensipleri:** Görselleştirmeleri bir bağlam içine oturtma, ana mesajı vurgulama, gereksiz detaylardan kaçınma, hedef kitleye uygun dil kullanma ve bir akış içinde sunma. Bu, teknik olmayan paydaşların bile veriden anlam çıkarmasını sağlar.

4.4 Veritabanı Yönetimi: SQL ve NoSQL

Veri bilimcilerin verileri depolamak, yönetmek ve sorgulamak için veritabanı sistemlerini anlamaları ve SQL kullanmaları kritik öneme sahiptir. Büyük veri ortamlarında NoSQL veritabanları da yaygınlaşmıştır.

- **İlişkisel Veritabanları:** Verilerin tablolar halinde düzenlendiği ve ilişkilerle bağlandığı sistemler (MySQL, PostgreSQL, Oracle, SQL Server). SQL ile veri çekme, birleştirme (JOIN), filtreleme (WHERE), sıralama (ORDER BY) ve gruplama (GROUP BY) işlemleri yapılır. Yapılandırılmış veriler için idealdir.
- **NoSQL Veritabanları:** İlişkisel olmayan, esnek şemalı veritabanları; büyük ve yapılandırılmamış veya yarı yapılandırılmış veriler için idealdir. MongoDB (belge tabanlı), Cassandra (sütun tabanlı), Redis (anahtar-değer) ve Neo4j (grafik tabanlı) örnek verilebilir. Ölçeklenebilirlik ve esneklik avantajı sunar.
- **Veri Ambarları ve Veri Gölleri:** Veri ambarları, analitik sorgular için optimize edilmiş yapılandırılmış veriyi depolarken; veri gölleri ham veriyi orijinal formatında depolayan merkezi alanlardır. Her ikisi de modern büyük veri mimarilerinin önemli bileşenleridir.



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



BÖLÜM 5: BÜYÜK VERİ MİMARİLERİ

Büyük veri; geleneksel veri işleme yazılımlarının depolama, yönetme ve analiz etme kapasitesini aşan, **hacim, hız ve çeşitlilik (3V)** özelliklerine sahip veri setlerini ifade eder. Modern veri bilimi projeleri genellikle bulut tabanlı büyük veri mimarileri üzerinde çalışır; bu da ölçeklenebilirlik, esneklik ve maliyet etkinliği sağlar.

5.1 Büyük Veri Ekosistemi ve Bulut Platformları

Kavram	Açıklama	Örnek Teknoloji
Veri Gölleri	Ham veriyi orijinal formatında, herhangi bir şema veya yapıya bağlı kalmadan depolayan merkezi depolama alanları.	AWS S3, Azure Data Lake Storage, Google Cloud Storage
Veri Ambarları	Yapılandırılmış ve temizlenmiş veriyi, iş zekası ve analitik sorgular için optimize edilmiş şekilde depolayan sistemler.	Amazon Redshift, Google BigQuery, Snowflake
Hadoop Ekosistemi	Büyük veri setlerini dağıtık biçimde depolayan ve işleyen açık kaynaklı çerçeve.	HDFS, MapReduce, Hive, Pig, HBase
Apache Spark	Büyük veri işleme için hızlı, genel amaçlı ve dağıtık küme bilgi işlem sistemi.	Spark SQL, Spark Streaming, MLlib, GraphX
Bulut Veri Platformları	Büyük veri depolama, işleme, analitik ve ML hizmetleri sunan entegre altyapılar.	AWS Sagemaker, Azure Databricks, Google Vertex AI
Veri Akışı İşleme	Gerçek zamanlı veya yakın gerçek zamanlı olarak sürekli akan veri akışlarını işleme yeteneği.	Apache Kafka, Apache Flink, AWS Kinesis

5.2 Veri Entegrasyonu: ETL ve ELT

Veri bilimi projelerinde farklı kaynaklardan gelen verilerin bir araya getirilmesi ve analize hazır hale getirilmesi kritik öneme sahiptir. Bu süreçler genellikle ETL veya ELT yaklaşımlarıyla yönetilir.

- **ETL (Extract, Transform, Load):** Verilerin kaynak sistemlerden çıkarılması, analiz gereksinimlerine uygun olarak dönüştürülmesi ve hedef sisteme (veri ambarı) yüklenmesi sürecidir. Geleneksel veri ambarı yaklaşımlarında yaygın kullanılır.
- **ELT (Extract, Load, Transform):** Verilerin doğrudan hedef sisteme (veri gölü veya bulut veri ambarı) yüklenmesi ve dönüştürme işlemlerinin hedef sistem içinde yapılması



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



sürecidir. Büyük veri ve bulut ortamlarında daha esnek ve ölçeklenebilir olduđu için tercih edilir.



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



BÖLÜM 6: PROJE YÖNETİMİ VE ETİK

Veri bilimi projeleri sadece teknik becerileri değil, aynı zamanda sağlam proje yönetimi, operasyonel mükemmellik (MLOps) ve etik konularında hassasiyet gerektirir. Bu unsurlar projelerin başarılı tamamlanmasını ve topluma faydalı olmasını sağlar.

6.1 MLOps ile Süreç Yönetimi

MLOps (Makine Öğrenmesi Operasyonları), makine öğrenmesi modellerinin yaşam döngüsünü (geliştirme, dağıtım, izleme, bakım) otomatikleştirmeyi ve standartlaştırmayı amaçlayan bir dizi uygulama ve prensiptir. DevOps prensiplerinin makine öğrenmesi projelerine uygulanmasıdır. MLOps, modelin üretim ortamında güvenilir, ölçeklenebilir ve sürdürülebilir biçimde çalışmasını sağlar.

- **Otomatikleştirilmiş Veri Boru Hatları:** Veri alımı, temizliği, özellik mühendisliği ve model eğitimi süreçlerinin otomatikleştirilmesi. Bu, veri kalitesini ve modelin güncelliğini sağlar.
- **Model Versiyonlama ve Yönetimi:** Farklı model versiyonlarının takibi, depolanması ve yönetimi. Modellerin izlenebilirliğini ve gerektiğinde geri alınabilirliğini sağlar.
- **Sürekli Entegrasyon/Sürekli Dağıtım (CI/CD):** Model kodunun ve bağımlılıklarının otomatik olarak test edilmesi, entegre edilmesi ve dağıtılması. Geliştirme süreçlerini hızlandırır ve hataları azaltır.
- **Model İzleme:** Dağıtılan modellerin performansının (doğruluk, gecikme, veri kayması, model eskimesi) gerçek zamanlı izlenmesi. Performans düşüşleri tespit edildiğinde otomatik uyarılar ve müdahaleler sağlanır.
- **Model Yeniden Eğitimi:** Veri dağılımındaki değişiklikler veya performans düşüşleri durumunda modellerin otomatik veya manuel olarak yeniden eğitilmesi ve güncellenmesi.

6.2 Veri Etiği ve Sorumlu Yapay Zeka

Veri bilimi projelerinin toplumsal kabulü ve sürdürülebilirliği için etik ilkelere bağlılık esastır. Genç kadınların bu alandaki farkındalığı ve liderliği, geleceğin yapay zeka sistemlerinin daha adil, şeffaf ve insan odaklı olmasını sağlayacaktır.

- **Veri Gizliliği ve Güvenliği:** Kişisel verilerin korunması, anonimleştirme, şifreleme teknikleri ve GDPR ile KVKK gibi düzenlemelere tam uyum. Diferansiyel gizlilik gibi ileri teknikler, veri setlerindeki bireysel bilgilerin açığa çıkmasını engeller.
- **Algoritmik Tarafsızlık ve Adalet:** Veri setlerindeki veya algoritmik modellerdeki önyargıların (bias) tespiti, ölçülmesi ve azaltılması. Farklı demografik gruplar için adil ve eşit sonuçlar üretme çabası. Örneğin kredi başvurularında veya işe alım süreçlerinde algoritmik ayrımcılığın önlenmesi.



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



- **Şeffaflık ve Açıklanabilirlik (XAI):** Analiz süreçlerinin ve model kararlarının anlaşılabilir olması; özellikle kritik kararlar veren modellerin nasıl çalıştığının açıklanabilmesi. Kara kutu modellerinden kaçınmak güveni artırır.
- **Hesap Verebilirlik:** Veri bilimi projelerinin ve yapay zeka sistemlerinin neden olduđu sonuçlardan kimin sorumlu olduđunun belirlenmesi ve bu sorumluluğun şeffaf biçimde üstlenilmesi.
- **İnsan Odaklılık ve Denetim:** Yapay zeka sistemlerinin insan refahına hizmet etmesi, insan özerkliğini desteklemesi ve kritik kararlarda her zaman insan denetiminin sağlanması. Yapay zeka, insan kararını destekleyen bir araç olmalı; onun yerini almamalıdır.

Sorumlu Veri Bilimi: Temel Sorular

Bu modelin verisi kimleri temsil ediyor; kimleri dışarıda bırakıyor?

Algoritmanın hatası kimin için daha yüksek maliyetli olabilir?

Modelin kararını sıradan bir kullanıcıya açıklayabilir miyim?

Verilerimi nasıl topladım; bu süreçte rıza ve şeffaflık sağlandı mı?

Bu model uygulamaya alındığında hangi grupları olumsuz etkileyebilir?



Bu proje Avrupa Birliği tarafından finanse edilmektedir.



BÖLÜM 7: KARIYER YOLLARI VE GELECEK

7.1 Veri Bilimi Kariyer Olanakları

Veri bilimi alanı, sürekli büyüyen ve gelişen bir ekosistem sunmaktadır. Genç kadınlar için bu alanda farklı uzmanlık ve ilgi alanlarına yanıt veren çok sayıda kariyer fırsatı bulunmaktadır.

Kariyer Yolu	Temel Sorumluluklar	Aranan Yetkinlikler
Veri Bilimci	Veri analizi yapar, makine öğrenmesi modelleri geliştirir, iş problemlerine veri odaklı çözümler sunar.	İstatistik, programlama, alan uzmanlığı, iletişim.
Veri Mühendisi	Veri altyapılarını tasarlar, kurar ve yönetir; temiz ve erişilebilir veri sağlar.	SQL, ETL/ELT, bulut platformları, dağıtık sistemler.
Veri Analisti	Veriyi raporlar, görselleştirir, iş zekası araçlarıyla içgörü üretir.	SQL, Excel, Tableau/Power BI, iş analizi.
MLOps Mühendisi	Makine öğrenmesi modellerinin yaşam döngüsünü otomatikleştiren altyapıları kurar.	DevOps, konteynerizasyon, CI/CD, bulut servisleri.
İş Zekası Uzmanı	Karar destek panoları ve raporlar üretir, veriyi yöneticiler için anlamlı hale getirir.	Veri ambarı, görselleştirme, iş süreçleri bilgisi.

7.2 Geleceğin Trendleri

Veri dünyası sürekli evrim geçirmektedir. Bu alanda başarılı olmak için sadece mevcut araçları öğrenmek yetmez; geleceğin trendlerini anlamak ve bunlara hazırlıklı olmak da gerekir.

- Üretken Yapay Zeka ve Büyük Dil Modelleri (LLM):** Veri analizi süreçlerinde LLM'lerin kullanımı yaygınlaşmaktadır. Veri ön işleme, kod üretimi, doğal dilde sorgulama ve rapor üretimi gibi alanlarda dönüştürücü etkileri vardır.
- Küçük Dil Modelleri (SLM):** Daha az kaynakla çalışabilen, daha hızlı ve maliyet etkin modellerin geliştirilmesi; yapay zekanın daha geniş kitlelere ulaşmasını sağlamaktadır.
- Multimodal Yapay Zeka:** Metin, görüntü, ses ve video gibi farklı veri türlerini aynı anda işleyebilen sistemlerin yükselişi; veri bilimi uygulamalarının kapsamını genişletmektedir.
- Sentetik Veri Üretimi:** Gerçek veri yerine sentetik olarak üretilmiş verilerin model eğitimi ve test süreçlerinde kullanılması; veri gizliliği endişelerini azaltır ve veri toplama maliyetlerini düşürür.
- Sorumlu Yapay Zeka ve Yönetişim:** Etik, adil, şeffaf ve güvenli yapay zeka geliştirme yaklaşımları; düzenleyici çerçevelerin (EU AI Act gibi) yaygınlaşmasıyla birlikte daha kritik hale gelmektedir.



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



- **Uç Yapay Zeka (Edge AI):** Modellerin bulut yerine doğrudan cihazlar üzerinde çalıştırılması; gecikmeyi azaltır, gizliliđi artırır ve internet bağlantısı olmayan ortamlarda yapay zeka kullanımını mümkün kılar.



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



SONUÇ: VERİ ODAKLI LİDERLİK VE TOPLUMSAL ETKİ

Veri bilimi ve analitiđi, sadece bir meslek alanı olmanın ötesinde, dünyayı anlama, dönüştürme ve geleceđi şekillendirme biçimidir. Bu el kitapçığı, sizlere bu heyecan verici alanda sağlam, güncel ve stratejik bir başlangıç noktası sunmayı amaçlamıştır. KızBaşına Genç Yetenek Geliştirme Programı mezunları olarak edindiđiniz yetkinliklerle hem kendi kariyerlerinizde fark yaratacak hem de veri odaklı kararlarla daha iyi, daha adil ve daha sürdürülebilir bir gelecek inşa etme potansiyeline sahip olacaksınız.

Veri biliminin gerçek deđeri, üretilen modellerin teknik dođruluđunun ötesinde, bu modellerin gerçek bir soruna gerçek bir çözüm üretip üretmediğindedir. Bir belediyenin hizmet planlamasını iyileştiren, bir kadın güvenliđi haritasını besleyen, bir afet hazırlık politikasını yönlendiren ya da bir sosyal hizmetin daha dođru hedeflenmesini sađlayan bir analiz; teknik mükemmellikten çok daha fazlasını ifade eder.

Unutmayın, veri dünyası sürekli evrim geçirmektedir. Sürekli öğrenmeye, yeni araçları, teknikleri ve etik yaklaşımları keşfetmeye devam edin. Büyük veri, bulut bilişim, yapay zeka ve makine öğrenmesi alanındaki entegrasyonlar; veri biliminin geleceđini şekillendirecektir. Özellikle küçük dil modelleri, multimodal yapay zeka ve sentetik veri üretimi gibi alanlar, veri biliminin uygulama yelpazesini daha da genişletecektir.

Bu kitapçığı eline alan her okuyucu için aynı koşullar; aynı kaynaklar ya da aynı destek mekanizmaları geçerli olmayacaktır. Bu durum, harekete geçmeyi ertelemenin deđil; mevcut koşullara uygun bir başlangıç noktası belirlemenin gerekçesidir. Küçük ölçekli bir proje, açık kaynak bir veri seti üzerinde yapılan bir analiz ya da bir kamu kurumuna iletilen veri temelli bir politika önerisi; büyük dönüşümlerin tohumudur.

Verinin gücünü kullanarak toplumsal sorunlara çözüm üretme ve pozitif deđişim yaratma potansiyeliniz sınırsızdır. Geriye kalan, bu potansiyeli gerçek bir soruna, gerçek bir veriye ve gerçek bir ekiple uygulamaktır.



Bu proje Avrupa Birliđi tarafından finanse edilmektedir.



REFERANSLAR

- [1] Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- [2] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.
- [3] McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
- [4] IBM. (2024). *What is Data Science?* <https://www.ibm.com/cloud/learn/data-science>
- [5] CRISP-DM. (2024). *Cross-Industry Standard Process for Data Mining*. <https://www.crisp-dm.org/>
- [6] KızBaşına Genç Yetenek Geliştirme Programı İzleme ve Deđerlendirme Teknik Final Raporu (2026).
- [7] Zou, H., & Hastie, T. (2005). *Regularization and variable selection via the elastic net*. Journal of the Royal Statistical Society: Series B, 67(2), 301–320.
- [8] AWS. (2024). *What is a Data Lake?* <https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/>
- [9] Google Cloud. (2024). *What is MLOps?* <https://cloud.google.com/resources/mlops-whitepaper>
- [10] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- [11] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- [12] Chen, M., Mao, S., & Liu, Y. (2014). *Big Data: A Survey*. Mobile Networks and Applications, 19(2), 171–209.
- [13] European Commission. (2019). *Ethics Guidelines for Trustworthy AI*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- [14] GDPR. (2016). *General Data Protection Regulation*. <https://gdpr-info.eu/>
- [15] KVKK. (2016). *Kişisel Verilerin Korunması Kanunu*. <https://www.kvkk.gov.tr/>

Genç Yetenek Geliştirme Programı, KızBaşına tarafından Ankara Büyükşehir Belediyesi ortaklığıyla yürütölmekte olup Sivil Katılım Projesi kapsamında Avrupa Birliđi tarafından finanse edilmektedir. Proje, Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından Türkiye Belediyeler Birliđi ortaklığında, Dışişleri Bakanlığı Avrupa Birliđi Başkanlığı işbirliğinde yürütölen Sivil Katılım Hibe Programı kapsamında hayata geçirilmektedir. Bu yayın Avrupa Birliđi'nin mali desteđiyle Sivil Katılım Hibe Programı kapsamında hazırlanmıştır. Proje kapsamındaki tüm içerikler yalnızca KızBaşına sorumluluğundadır; UNDP ve Avrupa Birliđi'nin görüşlerini yansıtmamaktadır.