

# Veri Bilim - Yapay Öğrenme Yaz Okulu, 2017

## “Matematiksel Temeller ve Vaka Çalışmaları”

Boğaziçi Üniversitesi, [TETAM](#), Kandilli, İstanbul

### Konu ve Kapsam

Bu yaz okulunda veri bilim ve yapay öğrenme alanında karşılaşılan temel konular hem kuramsal, hem de uygulamalı açıdan ele alınacaktır. Yaz okulunu tamamlayan katılımcılara, yapay öğrenmede kullanılan matematiksel yöntemlerin temelleri, hesaplama araçları ve uygulamalar konusunda temel bir altyapı kazandırılması amaçlanmaktadır. Daha önce düzenlediğimiz matematiksel temeller odaklı iki okulun bilgisine hazırladığımız [sayfadan](#) ulaşabilirsiniz.

### Hedef Kitle

Üniversite öğrencileri, akademisyenler ile yapay öğrenme konusuna ilgi duyan profesyoneller yaz okulunun hedef kitesini oluşturmaktadırlar. Dersleri takip etmek için temel bir analiz, lineer cebir ve olasılık bilgisi ile başlangıç seviyesinde programlama (özellikle Python) becerisi yeterli olacaktır.

Üniversite öğrencileri ve akademisyenler için yaz okulu ücretsizdir. Profesyoneller içinse ücret 240 TL olarak belirlenmiştir. Ödemeyi, yaz okuluna geldiğiniz ilk gün yapabilirsiniz. TETAM masrafları düşüldükten sonra toplanan para Çağdaş Yaşamı Destekleme Derneği'ne bağışlanacaktır.

Yaz okuluna **kayıt** olmak için lütfen [formu](#) doldurunuz.

### Program

	23 Eylül Cumartesi	24 Eylül Pazar		
9:00-12:00	Veri Bilim Yöntemlerine Genel Bakış (Taylan Cemgil)	Optimizasyon (İlker Birbil)		
14:00-17:00	Veri Bilim Uygulamalarına Genel Bakış (Taylan Cemgil)	PARALEL OTURUMLAR		
17:00-18:00	PANEL	Vaka Çalışmaları 1 Güdümlü Öğrenme (Birol Yüceoğlu)	Optimizasyon Uygulamaları (Kaan Öztürk)	
	30 Eylül Cumartesi	1 Ekim Pazar		
9:00-12:00	Simülasyon (Sinan Yıldırım)	Paralel Programlama (Kamer Kaya)		
14:00-18:00	PARALEL OTURUMLAR		PARALEL OTURUMLAR	
	Vaka Çalışmaları 2 Güdümsüz Öğrenme (Birol Yüceoğlu)	Simülasyon Uygulamaları (Kaan Öztürk)	Paralel Öğrenme (Kamer Kaya)	Hadoop Üzerinde Büyük Veri İşleme Uygulamaları (Ahmet Demirelli)

## Kullanılacak Programlar ve Veri Kümeleri

Yaz okulunda kullanılacak programlar ve veri kümelerine GitHub [sayfasından](#) ulaşabilirsiniz.

## Ders İçerikleri

### Veri Bilim Yöntemlerine ve Uygulamalarına Genel Bakış (Taylan Cemgil)

1. Giriş ve tarihçe, yapay zeka, makine öğrenmesi, veri madenciliği
2. Güdümlü ve güdümsüz öğrenme
3. Regresyon, yapay sinir ağları
4. Bayesci istatistik ve olasılık modelleri
5. Bayesci çıkarım
6. Uygulama alanlarından örnekler
7. Veri analizine model tabanlı ve algoritma tabanlı yaklaşımlar

### Optimizasyon (İlker Birbil)

1. Genel çerçeve ve modelleme
2. Doğrusal ve tamsayılı programlama ve örnekler
3. Doğrusal olmayan programlama
4. Yapay öğrenme problemlerinde çözüm yöntemleri

### Vaka Çalışmaları 1 - Güdümlü Öğrenme (Biol Yüceoğlu)

1. Kredi risk tahmini
  - a. Verinin temizlenmesi
  - b. Öznitelik türetme
  - c. Tahmin modellerinin uygulanması
  - d. Parametrelerin belirlenmesi
2. Google Street View fotoğraflarından karakter tanıma
  - a. Veri kümelerinin oluşturulması
  - b. Tahmin modellerinin uygulanması
  - c. Veri kümesinin zenginleştirilmesi

### Optimizasyon Uygulamaları (Kaan Öztürk)

1. En dik iniş, Newton, ve sahte-Newton algoritmaları
2. Stokastik dik iniş, mini yığın algoritmaları
3. Çeşitli optimizasyon algoritmalarıyla el yazısı rakam tanıma
4. Scipy ile optimizasyon

### Simülasyon (Sinan Yıldırım)

1. Temel örnekleme yöntemleri
2. Önem örneklemesi
3. Markov zinciri Monte Carlo
4. Bayesci çıkarım uygulamaları

## **Vaka Çalışmaları 2 - GÜDÜMSÜZ ÖĞRENME (BİROL YÜCEOĞLU)**

1. Perakendede müşteri segmentasyonu
  - a. Veri kümesinin temizlenmesi
  - b. K-ortalamlar yöntemiyle segmentlerin belirlenmesi
  - c. Sonuçların değerlendirilmesi
2. MINST veri kümesi ile güdümsüz öğrenme

## **SİMÜLASYON UYGULAMALARI (KAAN ÖZTÜRK)**

1. PyMC3 ile Bayesci olasılık modelleri
2. Bir dağılımın parametrelerini kestirme
3. Doğrusal regresyon ve lojistik regresyon için parametre kestirimi
4. Değişim noktası modelleri
5. Saklı Markov modelleri

## **PARALEL PROGRAMLAMA (KAMER KAYA)**

1. Paralel ve dağıtık hesaplamaların temelleri
2. OpenMP ile çok çekirdekli işlemciler üzerinde paralel hesaplamaya giriş
3. CUDA ile GPU tabanlı paralel hesaplamaya giriş
4. Python ile paralel hesaplamaya giriş

## **PARALEL ÖĞRENME (KAMER KAYA)**

1. Python ile paralel makine öğrenmesi
2. Derin öğrenme

## **HADOOP ÜZERİNDE BÜYÜK VERİ İŞLEME UYGULAMALARI (AHMET DEMIRELLİ)**

1. Büyük veri depolama ve Hadoop
2. Hadoop yapısı ve bileşenleri
3. Hadoop komutları, Pig ve Hive kullanımı
4. Apache Spark işlerini Hadoop üzerinde çalıştırma
5. Spark ile RDD ve DataFrame yapıları
6. Spark ile veri dönüşümü işlemleri (Transformation ve Action metodları)
7. Apache Spark ML (Machine Learning) kütüphanesi

## **Dersleri Verenler**

### **İlker Birbil**

İlker Birbil, doktora çalışmasını 2002 yılında North Carolina State Üniversitesi'nde Endüstri Mühendisliği ana dalı ile Yöneylem Araştırması ve Matematik yan dallarında tamamlamıştır. 2002 - 2004 yılları arasında Erasmus Yönetim Araştırma Enstitüsü'nde (Hollanda) doktora sonrası araştırma bursu ile çalışmalarını sürdüren İlker Birbil, 2004 yılından bugüne dek Sabancı Üniversitesi'nde öğretim üyesi olarak görev almaktadır. Matematiksel programlama alanında algoritma tasarımı genel başlığı altında toplanabilecek kuramsal ve uygulamalı pek çok çalışması vardır.

### **Ali Taylan Cemgil**

Taylan Cemgil, Boğaziçi Üniversitesi bilgisayar mühendisliği bölümünde lisans ve yüksek lisans çalışmalarının ardından, yapay öğrenme konusundaki doktora çalışmalarını Hollanda, Nijmegen Radboud Üniversitesinde tamamladı. 2004 ve 2008 arasında doktora sonrası araştırmacı olarak sırasıyla Amsterdam ve Cambridge Üniversitelerinde çalıştı. 2008'den beri Boğaziçi Üniversitesi Bilgisayar mühendisliği bölümünde öğretim görevlisi olarak görev yapmaktadır. İlgili duyduğu araştırma alanları arasında Büyük verilerin analizi, Bayesci istatistiksel metotlar, yaklaşık çıkarım, yapay öğrenme konuları sayılabilir. Sanayi projelerinde veri analizi konularında danışmanlık çalışmalarının yanı sıra, Boğaziçi üniversitesinde konu ile ilgili yüksek lisans dersleri vermektedir.

### **Ahmet Demirelli**

Ahmet Demirelli, 2002 yılı Marmara Üniversitesi Matematik Öğretmenliği ve 2004 yılında Sabancı Üniversitesi Bilgi Teknolojileri Yüksek Lisans programından mezun oldu. 2003-2005 yılları arasında Sabancı Üniversitesi Mühendislik ve Doğa Bilimleri fakültesinde eğitim asistanı olarak görev yaptı. 2005 yılından bu yana Sabancı Üniversitesi BT yüksek lisans programında Java, kurumsal uygulama mimarileri, Linux sistem yönetimi ve mobil programlama dersleri vermektedir. BT yüksek lisans derslerinin yanında Veri Analitiği yüksek lisans programında Hadoop ile büyük veri işleme dersi vermektedir. İlgili alanları arasında büyük verilerin paralel olarak işlenmesi ve makine öğrenmesi uygulamaları sayılabilir.

### **Kamer Kaya**

Kamer Kaya, doktorasını 2009 yılında Bilkent Üniversitesi Bilgisayar Mühendisliği bölümünden almıştır. Daha sonra Fransa'daki CERFACS Araştırma Merkezi'nde Paralel Algoritmalar projesinde çalışmalarını sürdürmüş, 2011 yılında Ohio Eyalet Üniversitesi'ne doktora sonrası araştırma görevlisi olarak katılmıştır. 2012 yılında aynı üniversitede Yardımcı Doçent olarak çalışmaya başlayan Dr. Kaya, 2014 yılından beri Sabancı Üniversitesi'nde öğretim üyesi olarak görev almaktadır. Yüksek başarımlı hesaplama, paralel algoritmalar ve kriptografi alanlarında birçok çalışması bulunmaktadır.

### **Kaan Öztürk**

Kaan Öztürk, fizik lisans ve yüksek lisans derecelerini Boğaziçi Üniversitesi'nden, fizik doktora derecesini Rice Üniversitesi'nden aldı. Işık ve Yeditepe üniversitelerinin bilişim bölümlerinde öğretim üyesi olarak görev yaptı. 2015-2016 döneminde TÜBİTAK desteğiyle Rice Üniversitesi'nde ziyaretçi araştırmacı, 2016-2017 döneminde ise Sabancı Üniversitesi'nde doktoraüstü araştırmacı olarak çalıştı. Yapay öğrenme uygulamaları, hesaplamalı fizik, karmaşık sistemler konularıyla ilgileniyor.

**Sinan Yıldırım**

Sinan Yıldırım, Boğaziçi Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü'nde lisans ve yüksek lisans çalışmalarının ardından, matematiksel istatistik konusundaki doktora çalışmalarını İngiltere, Cambridge Üniversitesi'nde tamamladı. 2013 ve 2015 arasında doktora sonrası araştırmacı olarak İngiltere, Bristol Üniversitesi Matematik Bölümü'nde çalıştı. 2015'ten beri Sabancı Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi'nde öğretim görevlisi olarak görev yapmaktadır. İlgili duyduğu temel araştırma alanları Bayesci istatistik ve Monte Carlo yöntemleridir.

**Birol Yüceoğlu**

Birol Yüceoğlu, lisans ve yüksek lisans eğitimini Sabancı Üniversitesi Endüstri Mühendisliği bölümünde tamamladı. Doktora eğitimini 2015 yılında Maastricht Üniversitesi'nde Yöneylem Araştırması üzerine yaptı. Halen Migros T.A.Ş. Ar-Ge Merkezi'nde çalışmakta ve Sabancı Üniversitesi Veri Analitiği Tezsiz Yüksek Lisans programında ders vermektedir. Araştırma alanları veri analizi uygulamaları, çizge teorisi ve tamsayı programlamadır.