

Python ile Veri Bilimi –1 | Sınıf 7

Sertifika Programı | Kasım 2022

Program Ekibi

Akademisyen	Email
Prof. Dr. O. Örsan Özener Dr. Erinç Albey Dr. Pelin Atahan	orsan.ozener@ozyegin.edu.tr erinc.albey@ozyegin.edu.tr
Asistan Cem Demirel	Email cem.demirel@ozu.edu.tr
İdari Koordinasyon Vildan Şahin	Email vildan.sahin@ozyegin.edu.tr

Başlamadan Önce

Ön Koşullar

Python ile Veri Bilimi – 1 Sertifika Programı herhangi bir **mühendislik, temel bilimler veya benzeri sayısal disiplinden** lisans derecesine sahip veya **son sınıfta** olan kişilerin yarı zamanlı ve uzaktan katılımına uygun olarak tasarlandı. Ders materyallerini ve açık kaynak içerikleri takip edebilmek için iyi seviyede **İngilizce** bilgisi gereklidir. Python ve Pandas'a yeni olan veya uygulama tecrübesi sınırlı olan katılımcılar için derslerden önce kendi zamanlarında, program asistanının da desteğiyle tamamlayabilecekleri bir **hazırlık** programı da olacaktır.

Eğitim Hedefleri

Veri bilimi, çok farklı sektörlerde müşterileri anlama, geleceğe dair tahmin yapma ve öneriler geliştirmek gibi farklı amaçlar için kullanılıyor. Bu programda, veri biliminin bel kemiğini oluşturan analitik kavramlar, tahminleme modelleri, makina öğrenmesi ve model yerleştirme uygulamaları, matematiksel dayanaklarından başlanarak ve gerçek örnekler üzerinden hayata geçirilerek anlatılacaktır. Bu eğitimden sonra; finans, bilişim teknolojileri, üretim gibi çok farklı sektörler ile sosyal medya gibi alternatif veri kaynaklarından toplanan gerçek veriler ile ne gibi modeller kurulabileceğini ve bu modellerin gerçek sistemlerde nasıl devreye alındığını deneyimlemiş olacaksınız.

Program Tarihleri: 1 Kasım 2022 – 17 Nisan 2023

Canlı Ders Gün ve Saatleri: Pazartesi ve Çarşamba: 19:00 – 22:00

Program katılımcılarının senkron dersler, egzersiz ve ödevler için ayıracağı sürenin 160 saat ana program (+90 saat: Python bilmeyenler için program öncesi hazırlık çalışmaları) mertebesinde olacağını öngörüyoruz.

Derslere Katılım: Dersler **Zoom** üzerinden canlı / senkron ve asenkron olarak verilecektir. Tekrar izlemek veya acil durumlarda kaçırılan dersleri tamamlamak için tüm senkron dersler kaydedilecek ve program süresince erişime sunulacaktır.

Programı Tamamlama: Programı başarıyla tamamlayarak sertifika alabilmek için verilecek **haftalık egzersiz ve ödevlerin tümünün tamamlanması** ve **senkron derslerin en az %80'ine canlı olarak katılım sağlanması** gereklidir. Çevrimiçi derslere toplamda 24 x 3 saatlik ders zamanı ayırmanız beklenmektedir. Haftalık ödevler için asistan desteği sağlanacak ve geri bildirimler verilecektir.

* 90 saat hazırlık programı + 160 saat ana program olarak, senkron dersler, egzersiz ve ödevler için gereken süreler tahmin edilerek belirlenen ortalama öğrenci zamanıdır. Hazırlık içeriklerine ihtiyaç duymayan katılımcılar için program süresi ortalama 160 saattir. Gerçek zamanlar değişkenlik gösterebilir.

Program İçeriği

Hazırlık Modülü

Python Programlama ve Destekleyici Materyaller

Hazırlık materyalleri, program boyunca ihtiyaç duyacağınız **Programlama** ve **Temel Python** becerilerinizi tamamlamak ve pratik yapmak için **İngilizce** olarak hazırlanmış; hali hazırda internette yayınlanmış olan açık kaynaklı çeşitli ders videoları, alıştırma ve ödevlerden derlenmiştir. Python ile programlama ve veri analizi geçmiş olmayanların buradaki içerikleri program öncesinde kendi zamanlarında tamamlamalarını bekliyoruz. Program katılımcıları bu içerik ve egzersizleri tamamlarken asistan desteği alabilecektir. Python'a yeni olan veya programlama tecrübesi sınırlı olan katılımcıların bu egzersizlerle birlikte pratik düzeylerini yeterli seviyeye çekmelerini bekliyoruz.

Ders	Süre (Saat)	Ders İçeriği
Programming Foundations	12 Saat Ders + 45 Saat Çalışma	<ul style="list-style-type: none">o Preparing Python Environmentso What is Computation?o Branching and Iterationo String Manipulation, Guess and Check, Approximations, Bisectiono Decomposition, Abstractions, Functionso Tuples, Lists, Aliasing, Mutability, Cloningo Recursion, Dictionarieso Testing, Debugging, Exceptions, Assertionso Object Oriented Programmingo Python Classes and Inheritanceo Understanding Program Efficiency, Part 1o Understanding Program Efficiency, Part 2o Searching and Sorting

*Python
Foundations*

5 Saat Ders + 25 Saat Çalışma	<ul style="list-style-type: none"> o Numpy & Pandas Basics o Importing, Exporting Data o Indexing, Selecting, Assigning o Lambda Functions, Grouping, Mapping, Sorting o Merging, Joining, Appending, Concatenating o Visualization with Pandas and Matplotlib
--	--

Ana Modüller

Modül I – Mathematical Foundations of Data Science

<i>Gün</i>	<i>Zaman</i>	<i>Eğitim Türü</i>	<i>Ders İçeriği</i>
<i>23 Ocak Pazartesi</i>	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Analytics tool box: Optimization basics for analytics (concept of optimization, convexity, optimization in python)
<i>25 Ocak Çarşamba</i>	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Analytics tool box: Optimization basics implementation (use case regression from optimization perspective, ridge & lasso as constrained optimization), Lagrange relaxation
<i>30 Ocak Pazartesi</i>	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o The concept of variability and its importance o Probability axioms o Concept of independence, conditional probability, Bayes theorem, Naïve Bayes algorithm as a classifier
<i>1 Şubat Çarşamba</i>	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Concept of independence, conditional probability, Bayes theorem (ctd.) o Association Rule Mining & Market Basket Analysis
<i>6 Şubat Pazartesi</i>	19:00-20:00	Senkron / Canlı	Discussion of coding assignment / homework
<i>8 Şubat Çarşamba</i>	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Fundamental distributions (binomial, geometric, poisson, exponential, normal, student's t, chi-square, f-distribution, beta, gamma) and their use cases in predictive modeling (ctd.)
<i>9 Şubat Perşembe</i>	19:00-20:00	Senkron / Canlı	Discussion of coding assignment / homework

13 Şubat Pazartesi	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Introduction to evolution of random variables through time (stochastic processes – Markov Chains) o Monte Carlo simulations and Markov Chain Monte Carlo method
15 Şubat Çarşamba	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Fundamentals of frequentist and Bayesian statistics and their comparison
20 Şubat Pazartesi	19:00-22:00	Asenkron / Video	<ul style="list-style-type: none"> o Maximum likelihood estimation o Goodness of fit tests and curve fitting o Analysis of Variance with statistical perspective

Modul II – Principles of Learning from Data: Introduction to Machine Learning

Gün	Zaman	Eğitim Türü	Ders İçeriği
27 Şubat Pazartesi	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Introduction to machine learning (ML) o Supervised vs unsupervised learning o Overfitting/underfitting bias variance tradeoff o Metrics in machine learning o KNN
1 Mart Çarşamba	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Linear regression and extensions
6 Mart Pazartesi	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Linear regression and extensions
8 Mart Çarşamba	19:00-22:00	Asenkron / Video	<ul style="list-style-type: none"> o Regularization feature selection (ridge, lasso, others)
13 Mart Pazartesi	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Logistic regression and extensions
15 Mart Çarşamba	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Tree based method (decision trees)
16 Mart Perşembe	19:00-20:00	Senkron/ Canlı	Discussion of coding assignment / homework

<i>20 Mart Pazartesi</i>	19:00-22:00	Senkron / Canlı	o Ensemble methods (Random Forest, Bagging, Boosting, Xgboost, LightGBM)
<i>22 Mart Çarşamba</i>	19:00-22:00	Senkron / Canlı	o Ensemble methods (Random Forest, Bagging, Boosting, Xgboost, LightGBM)
<i>27 Mart Pazartesi</i>	19:00-22:00	Asenkron / Video	o Model validation: training, validation and test sets.
<i>29 Mart Çarşamba</i>	19:00-22:00	Senkron / Canlı	o Data issues (structured vs unstructured, missing data, outlier, unbalanced data duplicates) and remedies o Data similarity, clustering, k-means
<i>30 Mart Perşembe</i>	19:00-20:00	Senkron/ Canlı	Discussion of coding assignment / homework
<i>3 Nisan Pazartesi</i>	19:00-22:00	Asenkron / Video	o Support Vector Machines o Neural networks and extensions
<i>5 Nisan Çarşamba</i>	19:00-22:00	Senkron / Canlı	o Neural networks and extensions

Modul III – Data Science at Scale: Big Data Project Deployment in the Cloud

<i>Gün</i>	<i>Zaman</i>	<i>Eğitim Türü</i>	<i>Ders İçeriği</i>
<i>10 Nisan Pazartesi</i>	19:00-22:00	Senkron / Canlı	<ul style="list-style-type: none"> o Introduction to Information and Data Management Architectures o Conceptual Introduction to Cloud (AWS, GCP, Azure, Oracle) platforms o Basics of big data technologies (like Hadoop, Kafka) required to process, store and retrieve live stream data
<i>12 Nisan Çarşamba</i>	19:00-22:00	Asenkron / Video	<ul style="list-style-type: none"> o Hands on application for deployment of a real-life analytics project <ul style="list-style-type: none"> o Deploying models as a web service o Serving models as a web page o Dockerization – deploying models using docker containers o Usage of APIs (such as twitter API) to stream live data
<i>13 Nisan Perşembe</i>	19:00-20:00	Senkron/ Canlı	Discussion of coding assignment / homework
<i>17 Nisan Pazartesi</i>	19:00-22:00	Asenkron / Video	o Hands on application for deployment (ctd.)