



BRG Eğitim ve Danışmanlık BİLGİ Veri Bilimi ve Optimizasyon Araştırma Grubu

Veri Bilimi Eğitim Programı

Bu programın hedefi katılımcılarına veri güdümlü bir bakış açısı kazandırarak, veri biliminin temel kavramlarını ve modern araçlarını çeşitli uygulamalar üzerinde çalışarak öğrendikleri bir deneyim sunmaktır. Altı farklı ders ile veri biliminin temel alanlarını kapsayan programda her bir ders dört ayrı günde üçer saatlik çevrimiçi buluşmalarla aktarılacaktır. Tüm dersler sigortacılık, finans veya ilgili bir alandan seçilen büyük veri setleriyle motive edilmekte ve her dersin sonunda somut çıktılar alınmaktadır. Derslerde Python programlama dili PyCharm arayüzü ile kullanılacaktır. Ders materyali, kod ve veri paylaşımı GitHub platformu üzerinden yapılacaktır. Katılımcıların Python dilini bilmeleri beklenmemektedir, ancak temel bir kodlama altyapısı faydalı olacaktır. Program başlamadan önce katılımcılara kısa bir video ders gönderilerek Python ve PyCharm kurulumu ile GitHub platformu hakkında bilgi verilecektir. Programı oluşturan dersler ve içerikleri aşağıdaki gibidir.

Video Ders – Python, PyCharm, GitHub

Program başlamadan önce tüm katılımcılarla paylaşılacak bu 20 dakikalık video derste Python ve PyCharm kurulumu ve arayüz tanıtımı yapılacak, GitHub platformu tanıtılacaktır.

Ders 1 - Python ile Veri Bilimine Giriş (Cemal Deniz Yenigün, Doruk Şen)

Bu dersi alan bir katılımcının aşağıdakilere benzer **örnek uygulamaları** ve daha fazlasını yapabilecek yetkinliğe gelmesi hedeflenmektedir.

- Ürün kategorilerine göre poliçe sayıları, gelir ve geri ödeme miktarlarının görselleştirilmesi.
- Sigorta taleplerinin çözümlenme sürelerinin hesaplanıp grafikler ve istatistiklerle özetlenmesi. En uzun ve en kısa süren sigorta ödeme işlemlerinin mercek altına alınması.
- Poliçe bazında tutulan büyük veriden il bazında en başarılı acentelerin belirlenmesi.
- Hasar kayıp oranlarının hesaplanması.

Python ile Veri Bilimine Giriş dersi aşağıdaki program çerçevesinde dört ayrı günde üçer saatlik çevrimiçi buluşmalarla aktarılacaktır.

Gün 1 – Veri Bilimine Giriş; Keşfedici Veri Analizi: Görsel ve Nümerik Veri Keşfi

Veri biliminin sigortacılık sektöründeki yeri ve geleceği. Sigortacılık sektöründen veri kullanarak basit grafiklerin (pie chart, bar chart, histogram, vb.) ve temel istatistiklerin elde edilmesi. Eldeki büyük verinin anlamlandırılması için ilk aşamada yapılması gereken görsel ve nümerik keşifler.

Gün 2 – Veri Tipleri, Yaygın Veri Formatları, Python'a Veri Aktarma

Kategorik ve nümerik veri tiplerinin özellikleri ve analiz şekilleri. Yaygın kullanılan .txt, .csv, .xlsx uzantılı veri setleri hakkında bilgi, bu veri setlerinin analiz için Python'a aktarılması.

Gün 3 – Veriyi Analize Hazırlama: İşleme, Dönüştürme, Özetleme

Analize hazır, düzenli veri (tidy data) kavramı. Verinin temizlenmesi, eksik veri için çözümler, filtreleme, sütunların düzenlenmesi, sütunların dönüştürülmesi, satırları sıralama, gruplama ve grup özetleri.

Gün 4 – Python ile Temel Programlama

Değişkenler, koşullu ifadeler, döngüler, vektörel operasyonlar üzerinden temel programlama becerisinin geliştirilmesi.

Ders 2 – Veri Görselleştirme (Umman Mahir Yıldırım)

Bu dersi alan bir katılımcının aşağıdakilere benzer **örnek uygulamaları** ve daha fazlasını yapabilecek yetkinliğe gelmesi hedeflenmektedir.

- Müşteri sayıları, poliçe sayıları, gelir ve geri ödeme miktarlarının yıllar içerisindeki değişiminin, varsa trend ve mevsimselliklerin görselleştirilmesi.
- Müşteri ya da acente kırılımında memnuniyet skorlarının farklı grafikler yardımıyla görselleştirilmesi.
- Sigorta taleplerinin sayıları, çözülme süreleri gibi belirlenen spesifik performans kriterlerinin il bazında harita üzerinde görselleştirilerek analiz edilmesi.
- Poliçe tipi bazında müşteri taleplerinin, kabul edilme oranları ve gelirlerin dağılımlarının görselleştirilmesi, odaklanılacak noktaların belirlenmesi.

Veri Görselleştirme dersi aşağıdaki program çerçevesinde dört ayrı günde üçer saatlik çevrimiçi buluşmalarla aktarılacaktır.

Gün 1 – Veri Görselleştirme Temelleri

Grafik türlerinin ve farklı senaryolar altında kullanım alanlarının tanıtılması, Python kütüphaneleri Matplotlib ve Seaborn ile veri görselleştirme.

Gün 2 – Gelişmiş Veri Görselleştirme Teknikleri

Coğrafi konum bilgisini kullanarak harita üzerinde veri görselleştirme, il/bölge bazlı analizler, ısı haritaları.

Gün 3 – İnteraktif Veri Görselleştirme

Kullanıcı tarafından belirlenmiş ve güncel veriye ya da kullanıcı seçimine göre değerleri dinamik olarak değişen belirli anahtar performans kriterlerinin (KPI), bütünsel bir bakış açısı ile tek ekranda adresleyecek kontrol panellerinin (dashboard) geliştirilmesi.

Gün 4 – Zaman Serisi Verileri

Müşteri sayıları, poliçe sayıları gibi zaman serisi olarak adreslenebilecek çok yıllık verilerde yıllar içerisindeki değişiminin, varsa anomalilerin, trend ve mevsimselliklerin çalışılması.

Ders 3 – İstatistiksel Öğrenme (Cemal Deniz Yenigün)

Bu dersi alan bir katılımcının aşağıdakilere benzer **örnek uygulamaları** ve daha fazlasını yapabilecek yetkinliğe gelmesi hedeflenmektedir.

- Sigortalıların yıllık sağlık giderleriyle kişisel özellikleri (yaş, cinsiyet, vücut kitle indeksi, eğitim düzeyi, vb.) arasındaki ilişkinin görselleştirilmesi, korelasyonların hesaplanması, korelasyon matrislerinin ve haritalarının oluşturulması. Bir regresyon çalışması yapılarak hangi kişisel özelliklerin sağlık harcamalarına ne kadar etki ettiğinin hesaplanması. Yeni bir müşterinin kişisel özelliklerine bakılarak yıllık sağlık giderinin tahmin edilmesi.
- Sigortalıların kaza yapıp yapmamalarıyla kişisel özellikleri (yaş, cinsiyet, sürüş deneyimi, önceki kazalar, eğitim düzeyi, araç tipi, kişisel/ticari araç vb.) arasındaki ilişkinin görselleştirilmesi. Bir sınıflandırma çalışması yapılarak hangi kişisel özelliklerin kaza ihtimalini arttırdığının belirlenmesi. Yeni bir müşterinin kişisel özelliklerine bakılarak kaza yapıp yapmayacağını tahmin edilmesi.
- Firmayı terk edecek müşterilerin önceden tahmin edilmesi için bir model kurulması. Farklı yöntemlerle kurulacak modellerin hata tiplerinin çapraz onaylama yöntemiyle karşılaştırılarak en az hatayla tahmin yapacak yöntemin benimsenmesi.

İstatistiksel Öğrenme dersi aşağıdaki program çerçevesinde dört ayrı günde üçer saatlik çevrimiçi buluşmalarla aktarılacaktır.

Gün 1 – Korelasyon Analizi, Basit Doğrusal Regresyon

Korelasyon, korelasyon matrisi, gözetimli öğrenme (supervised learning) ve gözetimsiz öğrenme (unsupervised learning) kavramları, basit doğrusal regresyon modeli, R^2 , regresyon ile tahmin, Python regresyon çıktılarının yorumlanması.

Gün 2 – Çoklu Doğrusal Regresyon

Birden çok tahmin edici ile doğrusal regresyon, doğrusal olmayan dönüşümlerle regresyon, kategorik tahmin ediciler (dummy variables), kademeli regresyonla (stepwise regression) değişken seçimi.

Gün 3 – Sınıflandırma (Classification) Problemleri

Sınıflandırma problemlerine giriş, lojistik regresyon, K-komşu modelleri (K-nearest neighbours), hata tiplerinin minimize edilmesi, ROC eğrisi.

Gün 4 – Çapraz Onaylama (Cross Validation) Yöntemleri

Eğitme verisi (training data) ve test verisi (test data), çapraz onaylama (cross validation), performans karşılaştırmaları ve en iyi tahmin yönteminin bulunması.

Ders 4 – Makine Öğrenmesi (Doruk Şen)

Bu dersi alan bir katılımcının aşağıdakilere benzer **örnek uygulamaları** ve daha fazlasını yapabilecek yetkinliğe gelmesi hedeflenmektedir.

- Mevcut sigortalıların kişisel özellikleri (yaş, cinsiyet, vücut kitle indeksi, eğitim düzeyi, vb.) ve poliçe kullanımları karakterize edilerek segmentlere ayrılması. Yeni müşterinin hangi segmente ait olduğunun belirlenmesi ve ücretlendirme politikasına yardımcı algoritmaların tasarlanması.
- Hayat sigortası için risk tahmini modelleri geliştirilmesi. Sigortalı tazminat talepleri için tahmin modelleri geliştirilerek finansman planlamasına destek sağlanması. Firma tarafından geliştirilen yeni bir ürünün hangi müşteriler tarafından benimsendiğinin tespit edilmesi. Model değerlendirme ve en iyi tahmin modellerinin tespiti.
- Call center ve web sitesi chatbot metinlerinin sınıflandırma ve aciliyet analizlerinin yapılması.

Makine Öğrenmesi dersi aşağıdaki program çerçevesinde dört ayrı günde üçer saatlik çevrimiçi buluşmalarla aktarılacaktır.

Gün 1 – Gözetimsiz Öğrenme (Unsupervised Learning)

Veri ön işleme, K-ortalamlar kümelemesi (K-means clustering), hiyerarşik kümeleme, müşteri segmentasyonu.

Gün 2 – Destek Vektör Makineleri (Support Vector Machines)

Destek Vektör Regresyonu (SVR), Destek Vektör Sınıflandırıcısı (SVC), çekirdek fonksiyonu seçimi ve parametre optimizasyonu.

Gün 3 – Yapay Sinir Ağları (Artificial Neural Networks)

Çok katmanlı algılayıcı (MLP), aktivasyon fonksiyonu, geri yayılım (Backpropagation), derin öğrenmeye giriş.

Gün 4 – Metin Madenciliği (Text Mining) ve Doğal Dil İşleme (NLP)

Metin ön işleme, metin görselleştirme, adlandırılmış varlık tanıma (NER), metin parçası (POS) etiketleme, metin sınıflandırma, kelime çantası (bag-of-words), konu modelleme.

Ders 5 – Özel Konular (Cemal Deniz Yenigün, Umman Mahir Yıldırım)

Bu dersi alan bir katılımcının aşağıdakilere benzer **örnek uygulamaları** ve daha fazlasını yapabilecek yetkinliğe gelmesi hedeflenmektedir.

- Markdown yazılımı ile yüksek kalitede teknik raporlar veya sunumlar hazırlanması.
- Veritabanındaki hiyerarşik verinin müşteri grubu, poliçe tipi, sigortalının demografik karakteristikleri gibi belirli kırılımlarda sorgulanarak hesap tablosu (spreadsheet) formatında çekilmesi.
- Veritabanında güncelleme işlemlerinin yapılması; yeni müşteri ekleme, bilgilerini güncelleme, poliçede tanıtılan yeni bir özelliği dikkate alarak ürün tablosunu güncelleme.

Özel Konular dersi aşağıdaki program çerçevesinde dört ayrı günde üçer saatlik çevrimiçi buluşmalarla aktarılacaktır.

Gün 1 – Markdown ile Teknik Rapor ve Sunum Hazırlanması

Markdown yazılımı ile Python kodunu birlikte kullanarak hesaplamaları ve dokümantasyonu birlikte yaparak yüksek kalitede teknik raporlar veya sunumlar hazırlamak. Word, PDF veya HTML formatlarında çıktı almak.

Gün 2 – İlişkisel Veritabanları ve SQL'e giriş

İlişkisel veritabanı tasarımı (tablolar ve normalizasyon), SQL'de veri tipleri, Veri Tanımlama Dili (DDL), Veri İşleme Dili (DML) ve Veri Sorgulama Dili (DQL), SELECT, DISTINCT, WHERE, CASE ifadeleri, filtreleme ve sıralama işlemleri.

Gün 3 – SQL'de Fonksiyonlar ve Tablo Sorguları

Tarih, metin fonksiyonları, matematiksel fonksiyonlar, tablo birleştirme, çoklu tablo sorgulama, toplama fonksiyonları (Aggregate Functions), GROUP BY, HAVING ifadeleri.

Gün 4 – Sektöre Özgü Vaka Çalışmaları

Müşteri terk tahminleri, sahtecilik ağlarının tespiti, risk değerlemeleri, ürün tavsiye sistemleri gibi konularda vaka çalışmaları.

Ders 6 – Proje Yönetimi ve Finansal Uygulamalar (Dr. Genco Fas - Prof. Dr. Cenktan Özyıldırım)

Gün 1 – Proje Yönetimi

Deterministik ve Rassal Proje Süreçleri - CPM ve PERT Teknikleri

Gün 2 – Proje Yönetimi

Oyun Teoretik Proje Yönetimine Bakış

Gün 3 – Finansal Uygulamalar

Sigortacılıkta Kullanılan Finansal Araçlar ve Fiyatlama Örnekleri

Gün 4 – Finansal Uygulamalar

Sigortacılıkta Kullanılan Finansal Araçlar ve Fiyatlama Örnekleri

BİLGİ Veri Bilimi ve Optimizasyon Araştırma Grubu (dsOpt)

Prof. Dr. Cemal Deniz Yenigün

Doç. Dr. Cemal Deniz Yenigün lisans ve yüksek lisans derecelerini Orta Doğu Teknik Üniversitesi'nden, doktora derecesini 2007 yılında Bowling Green State University'den İstatistik alanında almıştır. 2007-2014 yılları arasında Bilkent Üniversitesi İşletme Fakültesi'nde yardımcı doçent olarak görev alan Dr. Yenigün, 2014 yazında İstanbul Bilgi Üniversitesi Endüstri Mühendisliği Bölümü'ne katılmıştır ve aynı bölümde 2018 yılından beri Bölüm Başkanlığı görevini yürütmektedir.

Makaleleri Journal of Statistical Computation and Simulation, Computational Statistics and Data Analysis, Communications in Statistics ve Social Networks gibi saygın uluslararası dergilerde yayımlanan Dr. Yenigün'ün iki projesi TÜBİTAK tarafından desteklenmiştir. Çalışmaları istatistiksel öğrenme, regresyonda değişken seçimi, bağımlılık ölçütleri, yaşam analizi ve ağ analizi üzerinde yoğunlaşmıştır. Olasılık, istatistik ve büyük veri analizi üzerine dersler vermektedir.

Dr. Öğr. Üyesi Umman Mahir Yıldırım

Dr. Öğr. Üyesi U. Mahir Yıldırım lisans derecesini İstanbul Teknik Üniversitesi Endüstri Mühendisliği Bölümü'nden almış, Yüksek Lisans ve Doktora derecelerini ise Sabancı Üniversitesi'nde yine Endüstri Mühendisliği alanında tamamlamıştır. TÜBİTAK araştırma bursu ile 2014 ve 2015 yıllarında Southampton Üniversitesi ve Eindhoven Teknik Üniversitesi'nde araştırmacı olarak görev almasını takiben 2016 yılında İstanbul Bilgi Üniversitesi Endüstri Mühendisliği Bölümü'nde doktor öğretim üyesi olarak göreve başlamıştır.

Araştırma alanları Uygulamalı Optimizasyon, Makine Öğrenmesi ve Metasezgisel Yaklaşımlar üzerinde yoğunlaşmış olup makaleleri Computers & Industrial Engineering, Information Systems Frontiers, ve IEEE Transactions on Intelligent Transportation Systems gibi saygın uluslararası dergilerde yayımlanmıştır. Verdiği Endüstriyel Bilişim Sistemleri Tasarımı ve Eniyilemede Metasezgisel Yaklaşımlar gibi dersler yanında 10 yılı aşkın bir süredir de sektörde karar destek sistemleri, süreç otomasyonu, veri görselleştirme ve sayılarla ikna gibi konularda projeler yürütmekte ve eğitimler vermektedir.

Dr. Öğr. Üyesi Doruk Şen

Dr. Öğr. Üyesi Doruk Şen lisans derecesini Yeditepe Üniversitesi Endüstri ve Sistem Mühendisliği, Yüksek Lisans derecesini University of York'tan İşletme ve Finans bölümlerinden, Doktora derecesini ise 2020 yılında Marmara Üniversitesi Mühendislik Yönetimi alanlarından almıştır. Dijital Medya ve Danışmanlık sektörlerindeki tecrübelerinin ardından, İstanbul Bilgi Üniversitesi Endüstri Mühendisliği Bölümü'ne 2018 yılında Araştırma Görevlisi olarak katılmıştır. 2021 yılından itibaren aynı bölümde Dr. Öğr. Üyesi olarak görevini sürdürmektedir.

Makaleleri Information Systems Frontiers, Sage, Utilities Policy ve Energy gibi saygın uluslararası dergilerde yayımlanmış ve Makine Öğrenmesi alanında araştırmacılığını sürdürdüğü bir Bilimsel Araştırma Projesi bulunmaktadır. Araştırmaları Makine Öğrenmesi, Metasezgisel Yaklaşımlar ve FinTech alanlarında optimizasyon çalışmalarına yoğunlaşmıştır. Optimizasyon Teknikleri, Proje Yönetimi ve Finans Mühendisliği üzerine dersler vermektedir.