

R ve SQL Server Kullanarak Veri Bilimi

Learn via: **Classroom / Virtual Classroom / Online**

Duration: **4 Gün**

<https://bilginc.com/tr/egitim/r-ve-sql-server-kullanarak-veri-bilimi-4282-egitimi/>

Overview

R, istatistiksel analizler, veri madenciliği ve makine öğrenimi için en popüler ortam ve dillerden biridir. R'in yönetilen ve ölçeklenebilir sürümü SQL Server, Power BI ve Azure ML'de çalışır. Bu 4 günlük eğitimin ana konusu R dilidir. Bununla birlikte, eğitimde veri bilimi uygulamaları için MS BI paketinde bulunan Python, T-SQL, Power BI, Azure ML ve Excel gibi dillerin ve araçların nasıl kullanılacağı da gösterilmektedir. Laboratuvarlar R diline odaklanmaktadır; demolarda diğer dillerdeki kodlar da gösterilmektedir.

Eğitim Formatı:

Bu eğitimimiz, eğitmen sunumları ve laboratuvarlarda yapılacak bireysel çalışmalardan oluşmaktadır. Lab çalışmaları sırasında katılımcılar ağırlıklı olarak R dilini kullanmaktadır.

Her katılımcı, aşağıdaki yazılımların önceden yüklenmiş olduğu bir sanal makinede önceden hazırlanmış bir bilgisayarda çalışmaktadır:

ML Hizmetleri ile SQL Server 2017 veya 2019 Veritabanı Motoru (Veritabanı İçi)

AdventureWorksDW2017 demo veritabanı

Microsoft R İstemcisi

RStudio IDE

SQL Server Yönetim Stüdyosu

Prerequisites

Katılımcıların veri analizi ve SQL Server araçları hakkında temel bilgi sahibi olmaları gerekmektedir.

What You Will Learn

Bu eğitime katılanlar sıfırdan R ile programlamayı öğrenirler. Ücretsiz R motoru ve RStudio IDE kullanılarak temel R kodu da eğitimde tanıtılmaktadır. Bir veri bilimi projesinin yaşam döngüsü ayrıntılı olarak açıklanmaktadır. Katılımcılar, verilere genel bakışın nasıl gerçekleştirileceğini ve bir projedeki en az ilgi çekici görev olan veri hazırlamanın nasıl yapılacağını öğrenirler. Verilerin gözden geçirilmesi ve hazırlanmasından sonra analitik kısım, değişken çiftleri arasındaki ilişkileri analiz etmek için ara istatistiklerle başlar. Ardından eğitim, doğrusal bağımlılıkları araştırmak için daha gelişmiş yöntemler sunar.

Bir modelde çok fazla değişken olması başlı başına bir sorun yaratabilir. Eğitim, matris hesaplamalarının temelleri ile başlayarak özellik seçiminin nasıl yapılacağını göstermektedir. Daha sonra eğitim, denetimli ve denetimsiz öğrenme dahil olmak üzere daha gelişmiş veri madenciliği ve makine öğrenimi analizlerine geçmektedir. Eğitimde ayrıca tahmin, metin madenciliği ve takviyeli öğrenme gibi güncel konular da tanıtılmaktadır.

Son olarak katılımcılar, laboratuvarlar aracılığıyla SQL Server, Azure ML ve Power BI'da R kodunun nasıl kullanılacağını ve demolar aracılığıyla bahsedilen tüm araçların içinde Python'un nasıl kullanılacağını da öğrenmektedir.

Outline

Outline

Giriş bölümünü takiben modüller aşağıdaki gibi olacaktır:

1. Veri bilimi ve R'ye giriş

- İstatistik, veri madenciliği, makine öğrenimi...
- Veri bilimi projeleri ve yaşam süreleri
- R ile tanışın
- R aletleri

- R veri yapıları

Laboratuvar 1

1. Python ile tanışın

- Temel sözdizimi ve nesneler
- NumPy ve Pandas ile veri manipülasyonu
- Matplotlib ve seaborn kütüphaneleri ile görselleştirmeler
- Scikit-Learn ile veri bilimi

Tartışma: R vs Python

1. Verilere genel bakış

- Veri setleri, vakalar ve değişkenler
- Değişken türleri
- Kesikli değişkenler için giriş niteliğinde istatistik
- Sürekli değişkenler için tanımlayıcı istatistikler
- Temel grafikler
- Örneklem, güven düzeyi, güven aralığı

Laboratuvar 2

1. Veri hazırlama

- Türetilmiş değişkenler
- Eksik değerler ve aykırı değerler
- Düzleştirme ve normalleştirme
- Zaman serisi
- Eğitim ve test setleri

Laboratuvar 3

1. İki değişken arasındaki ilişkiler ve ilişkilerin görselleştirilmesi

- Kovaryans ve korelasyon
- Olasılık tabloları ve ki-kare testi
- T-testi ve varyans analizi
- Bayes çıkarımı
- Doğrusal modeller

Laboratuvar 4

1. Özellik seçimi ve matris işlemleri

- Doğrusal modellerde özellik seçimi
- Temel matris cebiri
- Temel bileşen analizi
- Keşfedici faktör analizi

Laboratuvar 5

1. Denetimsiz öğrenme

- Hiyerarşik kümeleme
- K-ortalamlar kümeleme
- Birliktelik kuralları

Laboratuvar 6

1. Gözetimli öğrenme

- Sinir Ağları

- Lojistik Regresyon
- Karar ve regresyon ağaçları
- Rastgele ormanlar
- Gradyan artırıcı ağaçlar
- K-en yakın komşular

Laboratuvar 7

1. Modern konular

- Destek vektör makineleri
- Zaman serisi
- Metin madenciliği
- Derin öğrenme
- Pekiştirmeli öğrenme

Laboratuvar 8

1. SQL Server ve MS BI'da R

- ML Hizmetleri (Veritabanı içi) yapısı
- SQL Server'da harici komut dosyalarını çalıştırma
- Bir modelin depolanması ve yerel tahminlerin gerçekleştirilmesi
- Azure ML ve Power BI'da R

Laboratuvar 9