

# R Programlama ile Büyük Verilerin Analizi Eğitimi

## Eğitim Hakkında

R Programlama ile Büyük Veriyi Analiz Etme eğitimi, katılımcılara bellek içi / disk üzerinde, H2O, Hadoop ve Apache Spark kullanarak dağıtılmış analizin nasıl kullanılacağını ve Microsoft Makine Öğrenimi Sunucusu ile R'nin nasıl entegre edileceğini öğretir.

## Neler Öğreneceksiniz

- Belleğe sığan büyük veri kümelerini işlemek için Data.table'ı kullanmayı,
- SQL sorguları aracılığıyla toplu işlemeyi anlamayı,
- R kullanarak çevrimiçi öğrenme stili modellerini uygulamayı,
- Hadoop ve Spark arasındaki farkı açıklamayı,
- HDFS dosya biçimini anlamayı,
- Bir R API aracılığıyla Spark'tan yararlanmak için SparkR paketini kullanmayı,
- Verileri H2O ve R API ile yönetmeyi,
- H2O ve R API kullanarak modelleri uygulamayı,
- Microsoft R Server ve Microsoft R Client'in nasıl çalıştığını açıklamayı,
- Farklı veri depolarında tutulan büyük verileri keşfetmek için R Client ile R Server'ı kullanmayı,
- Grafikleri ve grafikleri kullanarak verileri görselleştirmeyi,
- Büyük veri kümelerini dönüştürmeyi ve temizlemeyi,
- Büyük verilerden oluşturulan regresyon modellerini oluşturmayı ve değerlendirmeyi,
- Büyük veriden oluşturulan bölümlendirme modellerini oluşturmayı, puanlamayı ve devreye almayı
- SQL Server ve Hadoop ortamlarında R kullanmayı.

## Eğitim İçeriği

### Bellek İçi Büyük Veri: Data.table

Neden data.table'a ihtiyacımız var?

Neden o

Data.table'daki i ve j argümanları

Sütunları Yeniden Adlandırma

Yeni sütunlar eklemek

Binning verileri (sürekli ile kategorik)

Kategorik değerleri birleştirmek

Değişkenleri Dönüştürmek

Data.table ile gruplama işlevleri

Eksik verilerin ele alınması

Uzun - Geniş ve Geri

Veri kümelerini birleştirme

Veri kümelerini birlikte istifleme (birleştirme)

### **SQL Bağlantıları ve Sıralı veri güncellemeleri**

Çevrimiçi Öğrenmeyi Uygulama (Sıralı Model Güncellemeleri)

Biglm paketi

### **H2O Üzerinden Veri İşleme ve Makine Öğrenimi**

Ufuk20'ye Giriş

Kümeyi başlatma, durumu kontrol etme

Verileri İçer Aktarma, H2O'de manipülasyon

Yapılandırılmamış veri analizi: Word2Vec

H2O modellerinin montajı

Genelleştirilmiş Doğrusal Modeller

RandomForest

Gradyan Artırma Makinesi (GBM)

Topluluk model oluşturma

### **Hadoop'a Genel Bakış**

Dağıtılmış veriler ile dağıtılmış analitik karşılaştırması

HDFS ve harita azaltma

### **Apache Spark**

Spark'a Genel Bakış

Apache Spark'ı R ile kullanmak için API'ler

Sparklyr ve SparkR

R, Python, Java ve Scala API'lerinden Spark'a

SparkR Kullanılarak Uygulanan Örnekler

Spark (R) 'da veri içe aktarma ve işleme

Spark makine öğrenimi kitaplığı mllib:

Genel Doğrusal Modeller

### **Microsoft Machine Learning Sunucusuna Genel Bakış**

Microsoft R sunucusu nedir

Microsoft R istemcisini kullanma

ScaleR fonksiyonları

### **Veri Parçalama**

XDF dosyalarını anlama

Veri G / Ç

Değişken dönüşümler

Veri alt kümeleme, bölme ve birleştirme

### **Veri Özetleme**

Görselleştirmeler oluşturma

Sayısal özetler

### **Büyük Verilerin İşlenmesi**

Büyük Veriyi Dönüştürmek

Veri kümelerini yönetme

### **Genel Doğrusal Modellerin Uygulanması**

Bölümler / kümeler oluşturma ve bunlardan yararlanma

Regresyon modellerini uydurmak ve tahminlerde bulunmak