

Module 2 - Languages of Data Science

Bu notlar, veri biliminin temel yapı taşlarından biri olan programlama dillerinin kullanım alanları ve özellikleri hakkında rehber nitelikte düzenlenmiştir. Notların hazırlanması sırasında, [IBM Data Science Professional Certificate](#) adresindeki "Tools for Data Science" eğitiminin içeriği referans alınmış, ek olarak internet kaynakları ve yapay zeka araçlarının sunduğu bilgiler göz önünde bulundurulmuştur.

1. Dil Seçimi Kriterleri

Veri bilimi projelerinde kullanılacak programlama dilinin seçimi, projenin başarısında kritik bir rol oynadığı dikkate alınarak gerçekleştirilir. Bu seçimde göz önünde bulundurulan temel faktörler aşağıdaki gibidir:

- **Çözülecek Problem Türü:**

Veri analizi, web geliştirme, mobil uygulama veya sistem programlama gibi farklı alanların ihtiyaçlarına uygun diller tercih edilmektedir.

- **Ekibin Deneyimi:**

Projede görev alacak geliştiricilerin ve veri bilimcilerin hakim olduğu diller, işin hızlı ve verimli ilerlemesi açısından belirleyici kabul edilmektedir.

- **Performans Gereksinimi:**

Büyük veri ve gerçek zamanlı uygulamalarda düşük gecikme ve yüksek işlem hacmi destekleyen diller ön plana çıkarılmaktadır.

- **Ekosistem ve Kütüphaneler:**

Mevcut kütüphaneler, çerçeveler ve geniş topluluk desteği, dil seçiminde temel etmenler olarak değerlendirilmekte olup bu husus vurgulanmaktadır.

- **Hedef Platform:**

Web, masaüstü, mobil veya bulut tabanlı uygulama alanlarına göre farklı dillerin kullanımı tercih edilmektedir.

Popüler Diller: Python, R, SQL, Scala, Java, C++, Julia

Özel Durumlar İçin: JavaScript, PHP, Go, Ruby, Visual Basic

2. Python

Python, veri bilimi projelerinde en yaygın kullanılan dillerden biri olarak konumlandırılmakta ve çeşitli avantajları nedeniyle tercih edilmektedir.

- **Avantajları:**

- Açık ve okunabilir sözdizimi sayesinde öğrenme sürecinin hızlandırılması sağlanmaktadır.
- Geniş küresel topluluk desteği ve kapsamlı dokümantasyon sunulmaktadır.
- Bilimsel hesaplama ve yapay zeka uygulamaları için **Pandas, NumPy, SciPy, Matplotlib, TensorFlow ve PyTorch** gibi güçlü kütüphanelerin kullanılabilmesi mümkün kılınmaktadır.
- **NLTK** ile doğal dil işleme (NLP) uygulamalarında destek verilmektedir.

- **Büyük Kullanıcılar:**

IBM, Google, Facebook, NASA, Amazon gibi küresel ölçekte faaliyet gösteren kurumlar tarafından tercih edilmektedir.

- **Topluluk:**

PyLadies gibi topluluklar aracılığıyla çeşitlilik ve kapsayıcılık desteklenmekte, PyCon ve yerel etkinliklerle bilgi paylaşımının sürekliliği sağlanmaktadır.

Not: Python'un, makine öğrenimi, derin öğrenme ve veri analitiği alanlarında sunduğu kütüphane zenginliği nedeniyle veri bilimi projelerinde öncelikli dil olarak kullanıldığı vurgulanmaktadır.

3. R

R, özgür yazılım prensiplerine bağlı olarak geliştirilmekte olup istatistiksel analiz ve veri görselleştirme konularında güçlü yetenekler sunmaktadır.

- **Python ile Farklılıklar:**

R, "Free Software" (özgür yazılım) prensiplerine sıkı sıkıya bağlı olarak geliştirilmektedir; Python ise açık kaynak lisansı (ör. PSF License) kapsamında genel amaçlı bir dil olarak kullanılmaktadır.

- **Avantajları:**

- İstatistiksel analiz, veri madenciliği ve grafik üretimi konularında üstün performans sergilemesi sağlanmaktadır.
- Az veya hiç programlama deneyimi bulunan kullanıcıların bile hızlıca veri analizi yapabilmesi mümkün kılınmaktadır.
- **CRAN** üzerinde bulunan 15.000'den fazla paket ile sürekli genişleyen bir ekosistem sunulmaktadır.
- **C++**, **Java** veya **Python** ile entegrasyon imkânı sağlanarak esnek kullanım alanları oluşturulmaktadır.

- **Kullanıcılar:**

Akademik çevreler, araştırmacılar ve IBM, Google, Microsoft gibi büyük kurumlar tarafından tercih edilmektedir. Finans, sigorta ve farmasötik sektörlerde istatistiksel uygulamalar kapsamında kullanılmaktadır.

- **Topluluk:**

useR, **WhyR**, **R-ladies** ve **SatRdays** gibi topluluklar aracılığıyla bilgi paylaşımının ve öğrenmenin desteklendiği gözlemlenmektedir.

Not: R'ın, özellikle RStudio ile birlikte kullanıldığında, zengin görsel çıktı ve interaktif veri analizi süreçlerini desteklediği vurgulanmaktadır.

4. SQL

SQL, ilişkisel veritabanlarının yönetimi ve veri manipülasyonu için geliştirilmiş olup veri ile çalışan tüm rollerde kritik bir öneme sahip olarak değerlendirilmektedir.

- **Özellikleri:**

İlişkisel veritabanlarında veri sorgulama, ekleme, güncelleme ve silme işlemlerinin etkin bir şekilde gerçekleştirilmesi sağlanmaktadır.

- **Avantajları:**

- Tek bir veritabanında öğrenilen SQL bilgisinin, farklı veritabanı sistemlerinde uygulanabilir olması göz önünde bulundurulmaktadır.
- ANSI standardına uygun çeşitli sürümlerin (ANSI SQL, T-SQL, PL/SQL) bulunmasıyla esneklik sağlanmaktadır.

- **Desteklenen Veritabanları:**

MySQL, PostgreSQL, Oracle, SQLite, Microsoft SQL Server gibi sistemler ile Amazon RDS, Azure SQL Database gibi bulut tabanlı çözümler desteklenmektedir.

- **Uygulama Alanları:**

İş ve veri analistlerinin raporlama süreçlerinde, veri mühendisliği ve ETL uygulamalarında temel araç olarak kullanılmaktadır.

Not: SQL'in, veri manipülasyonu ve ilişkisel veritabanı yönetiminde sunduğu esneklik ve ölçeklenebilirlik nedeniyle kritik bir dil olarak öne çıkarıldığı vurgulanmaktadır.

5. Diğer Diller

Veri bilimi ekosisteminde, Python, R ve SQL'in yanı sıra belirli ihtiyaçlar doğrultusunda kullanılabilen diğer programlama dilleri de önemli roller üstlenmektedir.

1. Java:

- **Genel Amaçlı ve Hızlı:** Büyük kurumsal uygulamalarda, bankacılık sistemlerinde ve Android uygulama geliştirmede tercih edilmekte; yüksek performans ve ölçeklenebilirlik sağlanmaktadır.
- **Veri Bilimi Araçları:** **Weka, Java-ML, Deeplearning4j** ve **Spark MLlib** gibi araçların entegrasyonu ile veri analizi desteklenmektedir.

- **Büyük Veri Yönetimi:** Hadoop ekosisteminin bileşenlerinin Java ile geliştirilmiş olması, tercih sebeplerinden biri olarak değerlendirilmektedir.

2. Scala:

- **Java ile Tam Uyumluluk:** Java Sanal Makinesi (JVM) üzerinde çalışmakta ve fonksiyonel programlama paradigması sunulmaktadır.
- **Apache Spark:** Büyük veri ve akış verisi işleme senaryolarında tercih edilmekte; yüksek performanslı işlemler desteklenmektedir.
- **Performans ve Eşzamanlılık:** Yüksek performans ve eşzamanlı programlama yetenekleri ile öne çıkarılmaktadır.

3. C++:

- **Sistem Programlama ve Hız:** Düşük seviye işlemlerde yüksek performans ve etkin hafıza yönetimi sağlanmakta; altyapı projelerinde tercih edilmektedir.
- **Veri Bilimi Araçları:** Makine öğrenimi kütüphanelerinin çekirdeklerinin C++ ile geliştirilmiş olması, performans avantajı sunmaktadır.
- **Örnek Uygulamalar:** **MongoDB** ve **Caffe** gibi projelerde kullanımı desteklenmektedir.

4. JavaScript:

- **Web Tabanlı Geliştirme:** Tarayıcı ve sunucu tarafı uygulamalarda veri görselleştirme ve interaktif projeler kapsamında kullanılmakta; esnek yapısı ile dikkat çekmektedir.
- **TensorFlow.js:** Tarayıcı üzerinde makine öğrenimi modellerinin çalıştırılmasını mümkün kılmaktadır.

5. Julia:

- **Yüksek Performanslı Sayısal Analiz:** Bilimsel hesaplamalar ve büyük veri setleri ile çalışmada yüksek performans sunmakta; verimli paralel işlem yetenekleri desteklenmektedir.
- **JuliaDB:** Büyük veri setleriyle başa çıkma ve paralel işlem imkânı sunan özel veri tabanı çözümleri arasında değerlendirilmektedir.

- Akademik çevreler ve finans sektöründe tercih edilmeye başlanmakta, yükselen bir yıldız olarak konumlandırılmaktadır.
-

6. Özet

Veri bilimi projelerinde programlama dillerinin seçiminin, proje gereksinimleri, performans beklentileri ve ekibin deneyimi dikkate alınarak yapılmasının temel öneme sahip olduğu vurgulanmaktadır. Python ve R, geniş kütüphane desteği ve topluluk katkıları sayesinde veri analitiği ile makine öğrenimi projelerinde öncelikli olarak kullanılmaktadır. SQL ise ilişkisel veritabanı yönetimi ve veri manipülasyonu süreçlerinde kritik rol oynarken, Java, Scala, C++, JavaScript ve Julia gibi diller belirli senaryolarda ya da mevcut ekosistemle uyum sağlama amacıyla tercih edilmektedir. Doğru dil seçiminin, verimli, ölçeklenebilir ve sürdürülebilir çözümlerin oluşturulmasında temel bir etken olduğu dikkate alınmalıdır.