

Veri Bilimi Mühendisi (Yazılımcısı/Geliştiricisi)

Veri bilimi mühendisleri veri bilimi ekiplerinde veri kümelerinin toplanması, saklanması, ön işlenmesi, birleştirilmesi ve bu verileri kullanarak geliştirilen modellerin ya da algoritmaların ölçekli olarak canlıda kullanılır hale getirilmesi için yazılım geliştirmekle uğraşırlar.

Bu tanımı biraz daha açalım.

Veri bilimcilerinin veri analizi ve modellemesi yapabilmeleri ve bu analiz ve modellere dayanan iş süreci iyileştirme ya da karar destek sistemlerinin kullanılabilir hale getirilmesi için çeşitli yazılım ve veri otomasyon sistemleri inşa etmek gereklidir. Bunu sağlamak için Veri bilimi mühendislerinin üstüne düşen görevlere aşağıdaki örnekler verilebilir.

- veri bilimcileriyle birlikte çalışıp sistemin verilerinin toplanması için gereken veri akis sistemlerini tasarlamak ve geliştirmek
- geliştirilen modellerin canlıya (production) taşınması için gerekli yazılım geliştirme
- değişik metriklerin raporlanması ve veri panellerinde kullanılabilmesi için veri depoları tasarlamak ve gerçekleştirmek
- değişik veri kaynaklarındaki veri kümelerinin bir veri depolarına ya da veri tabanlarına kopyalanması için ayıklama, dönüştürme ve yükleme (ETL) sistemleri tasarlamak ve kurmak
- veri bilimcilerin geliştirdikleri model prototiplerini yazılım mühendisliği teknikleriyle dayanıklı ve ölçeklenebilir bir hale getirmek

Bu alandaki mühendisler genelde bilgisayar bilimi, bilgisayar, elektronik ya da çok daha çeşitli mühendislik alanlarından mezundurlar. Yazılıma yakın olması nedeniyle yazılım alanında olduğu gibi alaylı bir şekilde çok değişik eğitim alanlarından bu alana girip başarılı olanlara sıklıkla rastlanır. Çalışabilecekleri projelere aşağıdaki örnekler verilebilir.

- Bir tahminsel veri bilimcisinin geliştirdiği kredi skorlama modelini bir web sitesi ile entegre etmek ve onu binlerce ya da milyonlarca kullanıcıyı destekleyebilecek bir şekilde geliştirmek (Wells Fargo, Bank of America, Carvana, Vroom)
- Bir web sitesinde kullanıcıların aktivite olaylarını (event) veri akımı şeklinde depolayıp bunları diğer veri bilimcilerinin kullanımına hazır hale getirmek (Amazon, Google, Facebook)
- Bir optimizasyon modelinin kullanıcılar ya da başka sistemler tarafından kullanılmasını sağlamak için arayüz veya API geliştirmek (Amazon, Carvana, Walmart)
- Pazar hareketlerini modelleyen bir sistem için borsa verilerini gerçek zamanlı olarak sorgulayıp depolayan bir veri borulama sistemi geliştirmek (Morgan Stanley, RobinHood)

Aranan Özellikler:

Veri Bilimi Mühendisliği/Yazılımcısı aranan en önemli özellik veri ile ilgili değişik teknolojileri çok çabuk öğrenip ve kavrayıp bunları gerçek problemler üzerinde pragmatik olarak uygulayabilmektir. Bunu yaparken sadece yeni teknoloji kullanmış olmak için yapmamak bu teknolojilerin iyi ve kötü yanlarını çok iyi anlayıp probleme uygun teknolojileri seçebilmektir.

Bir baska genel özellik de bulut teknolojilerine hakimiyet ve bulut üzerindeki deęisik seçeneklerin artilarına ve eksilerine hakim olmaktır. Veri kümeleri buyudukce bu veri kümelerini işlemek için bulut sistemlerini kullanmak gerekmektedir.

Java, C++ gibi geleneksel programlama dillerine nazaran veri işlemeye daha uygun olan Python, Scala, Go gibi dillerde programlama tecrübesi daha çok aranır.

Zorlukları:

Veri Bilimi Mühendisliği/Yazılımcısı işlerinde en büyük zorluklar çoğunlukla veri bilimcileri ile ortak çalışma alanında ve bu çalışmada sınırların çiziminde ortaya çıkar. Veri Bilimi Mühendisliği/Yazılımcısının üstüne düşen veri bilimcisinin modellerini yaklaşımlarını özümseyip bunlara en uygun çözümleri seçmektir. Bu aşamada insan ilişkileri ve anlaşmazlık yönetimi (conflict management) önemli yetkinlikler olarak karsımıza çıkar.

Diğer bir zorluksa üssel bir hızda gelişen bilgi teknolojilerini takip edip öğrenmeye iş dışında da devam etmek gerekliliğidir. Bu nedenle bu tip işlere doğal olarak ilgi duymayan kişilere uygun olmayabilir.