

Video Metni

Bir Veri Bilimcisinin Hayatında Bir Gün

(Stephen Sherman): Daha önce, büyük bir kuruluşun parçası olarak bir öneri motoru oluşturdum ve sorunun farklı bölümlerine yönelik olarak her tür mühendis ve muhasebe kanalıyla çalıştım. Bu, en çok memnun olduğum sistemlerden biri, çünkü nihayet üst yöneticilerden mühendislere ve geliştiricilere, tüm seviyeler için anlaşılması kolay, oldukça basit bir çözüm buldum. Sonuç olarak, bu sistem oldukça karmaşık bir sistem kadar verimliydi ve üzerinde çok daha fazla vakit harcayabilirlerdi.

(Rafael B. Da Silva): Üniversitedeyken, yosunların çiçek açma zamanlarını tahmin etmemizi gerektiren bir sorun üzerinde çalışıyorduk. Bu yosun çiçekleri, su toksisitesinde bir yükselişe ve su arıtma şirketleri kanalıyla sorunlara neden olabiliyordu. Kimya mühendisliği altyapımızla bu tahmini yapamadık. Bu yüzden, çiçek açmanın ne zaman gerçekleşeceğini tahmin etmek için yapay sinir ağlarını kullanmaya başladık. Bu şekilde, su arıtma şirketleri bu sorunun üstesinden daha iyi gelebiliyordu.

(Murtaza Haider): Toronto'da, toplu taşıma Toronto Taşıma Komisyonu tarafından işletiliyor. Bu komisyona kısaca TTC diyoruz. Komisyon, bölgenin ve Kuzey Amerika'nın en önemli taşıma yetkililerinden biri. Bir gün, benimle iletişime geçtiler ve "Bir sorunuz var," dediler. Ben de, "Pekala, sorun nedir?" dedim. "Elimizde şikayet verileri var ve bunları analiz etmek istiyoruz, yardımınıza ihtiyacımız var," dediler. "Elbette, yardım etmekten mutluluk duyarım," dedim. "Kaç şikayetiniz var?" diye sordum. "Biraz," dediler. "Kaç tane?" dedim. Muhtemelen yarım milyon. "Eh, o halde üzerinde çalışmaya başlayalım," dedim.

Verileri aldım ve analiz etmeye başladım. Aslında, verileri tablo formatında tutma konusunda harika bir iş başarmışlardı; yapılandırılmamış veriler vardı. Tablo şeklindeki veriler şikayetin gelme vakti, alan kişi, şikayetin türü, çözülüp çözülmediği ve kimin hatası olduğu şeklindeydi. Yapılandırılmamış kısım ise, e-posta ve fakslerden oluşuyordu. Yarım milyon e-postaya baktığınızı ve cevaplar bulmaya çalıştığınızı düşünün.

Bilmek istediğim ilk şey, insanların neden şikayet ettiği, bu konuda bir örüntü olup olmadığı veya diğer günlere kıyasla daha fazla şikayetin olduğu bir gün olup olmadığıydı. Verilere bakıp tüm farklı formatlarda analiz etmiştim, ancak şikayetlerin belirli bir günde yüksek ve diğerlerinde düşük olduğuna dair bir işaret görememiştim. Ve bu durum bir ay kadar daha devam etti. Sonra bir gün, Toronto'da otobüsten iniyordum ve hala bunu düşünüyordum. Yere bakmadan indim ve bir su birikintisine bastım. Ayak bileğime kadar suya batmışım ve bir ayağım ıslak, diğeri kuruydu. Çok kızmışım. Geriye doğru yürürken bir anda anladım ve kendi kendime, "Dur bir saniye. Bugün beklenmedik bir yağmur yağdı ve bu duruma hazırlıksızdım. Bu yüzden ıslandım ve önüme bakmıyordum. Ya aşırı hava koşullarıyla TTC'nin aldığı şikayetler arasında bir ilişki varsa?"

Environment Canada web sitesine girip yağmur ve yağış miktarı, rüzgar ve gün ışığı verilerini aldım. Ve çok ilginç bir şey keşfettim. En fazla şikayet alınan 10 gün, insanların en çok şikayet ettiği on gün, havanın kötü olduğu günlerdi. Beklenmedik yağmur, hava sıcaklığındaki ani düşüş, çok fazla kar, oldukça rüzgarlı bir gün gibi. Bu yüzden TTC'nin üst yöneticilerine gittim ve "Bir iyi, bir de kötü haberim var," dedim. İyi haber; insanların belirli günlerde çok fazla şikayette bulunmalarının nedenini buldum. Bunun nedenini biliyorum. Kötü haber ise, bu konuda yapabileceğimiz hiçbir şey yok.