Veri Bilimi ve Endüstri Mühendisliği Üzerine Sohbetler -Kerem Demirtaş & Erdi Daşdemir

Veri bilimi üzerinden otonom araçların yönlendirilmesi ve yönetilmesi konusunu anlattı. Burada trafik akış modelleri, sensörlerin kontrolü, envanter problemlerinin çözümü ve simülasyon araçlarının kullanıldığını belirtti. Data kısmında optimizasyonun yapılması örnek olarak salgın hastalıkların ortaya çıkması, tarihsel dataların incelenmesi, potansiyel olarak ileride nerede ortaya çıkabileceğini tahminlemesinin yapıldığını anlatmıştır. Tez konusu olarak da ulaşım anlamında otomobiller üzerindeki sensörler vasıtasıyla topladığı datalar örnek olarak önünüzdeki araç saatte kaç kilometre hızla gidiyor bunun 360 derece belli bir uzaklığa kadar algılayabiliyor, 3 boyutlu görev verisi çıkararaki hangi şeritte kaç araç var bunun hesaplamasını yapıp, yaygınlaşan bulut sistemine aktarımı yapılıyor. Burada ölçeklenebilir sistem oluşturulması hedeflenmektedir. Bu anlamda Veri Biliminin hedefi sadece model inşa etmek değil, veriyi toplayıp, ön işleme yapıp, veriyi analiz edip, sonuçların paydaşlarla paylaşılması da önemlidir. Veri biliminin sonuç çıkarabilmesi için insan faktörünün her zaman olması gerekmektedir. Doğru soruları bulup doğru soruların çıkarılması ve doğru şekilde yorumlanması gerekmektedir. Burada dikkat edilmesi gereken bir nokta ise evet ben modeli yaptım fakat her zaman doğru sonuç vermeyebilir çünkü burada trend değişebilir, gerçek dünya ile kıyaslama yapmak, sorgulamak ve modeli güncellemek gerekmektedir. Bu nedenle çalıştığınız konu hakkında bilginiz yoksa sadece model kurarak sonuçları yorumlayamazsınız. Bu anlamda yapay zeka bir veri bilimi değil veri biliminin bir aracıdır. İnsan faktörü ve insan yargısı yine olacaktır.

Veri biliminde model kurulması ve sonuçların çıkarılması aşağıdaki adımlardan oluşmaktadır:

1. Problemin tanımlanması: Biz burada doğru problemi mi çözüyoruz. Örnek verecek olursak eğer 2. Dünya savaşı zamanında vurulan uçakların vurulma bölgeleri analiz edilerek vurulan kısımlarda güçlendirme yapılmış ve bunun sonucunda uçağı ağırlığı artmasına rağmen vurulan uçakların oranında bir değişiklik olmadığı görülmüştür. Fakat problem farklı bir noktadan yaklaşıldığında asıl sonucu değiştirecek kısımda sorulması gereken soru “ Dönmeyen uçaklar nerede?” ve “Dönmeyen uçaklar nereden vuruluyor” bu soruların cevabının analiz edilmesi sonucunda kokpit ya da motor kısımda vurulan uçakların dönemediği görülmüştür. Aslında güçlendir,lmesi gereken kısımlar, uçakların dönmesini engelleyen problemlerin oluştuğu yerdir.
2. Data Toplama ve Hazırlama: Burada iki tip data bulunuyor biri structured database olan yani kolonları olan, integer bulunan datalar. İkincisi ise unstructured olan text verisi olup belli imla hataları içeren datalardır. Bunların harmanlanıp bir füzyon verisi olarak kullanılması gerekmektedir.
3. EDA yani keşifsel veri analizi: Mean ya da medyanda zıplama neden kaynaklanıyor, bunu bulup buna göre data science modeli kurulması önemlidir. Burada verilebilecek örnek ise 1854 yılındaki Kolera salgının nedeni havadan bulaşma olarak görülürken John Snow isimli bilim adamı bunun nedenin sudan bulaşma olarak ileri sürmüştür. Bu nedeni de desteklemek için veri bilimini kullanmıştır. Ölüm raporlarını inceleyip adresleri toplayıp bakıldığında ölüm oranlarının adrese göre değiştiği görülmüştür. Bunun sonucunda Soho’da ortak bir kuyudan pompa ile su alınan yerin çevresinde ölüm oranının yüksek olduğu görülmüş olup koleranın buradan yayıldığını tespit etmişlerdir.
4. Model Oluşturma ve Algoritmalar

İç görüler çıkarma, sonuçları tahmin etme, matematiksel algoritmik modeller çıkarmayı içermektedir. Tanımlayıcı, tahminleyici, aksiyonların önerilmesinden oluşan süreçleri bulunmaktadır.

Burada Lewis Fry Richardson “Bütün modeller yanlış olacak, bazıları size bilgi verecek, bu modeller sonucunda siz belli çözümlere gideceksiniz.” demiştir. Lewis Fry Richardson diferansiyel denklemler kullanarak hava tahmini yapan bir model oluşturmuştur. Burada data kalitesi, oluşturulan modelin algoritmasının kalitesi data science kalitesi açısından önemlidir.

1. Evaluation: model kurulması burada tahmin ne kadar doğru metrikler ne durumdadır bunu değerlendirilmesi bulunmaktadır. Underfitting, overfitting ve well balanced modelleri bulunmaktadır. Burada önerilen dengeli model olup bazı noktaları görebiliyorsanız hiç görmediğiniz noktaları tahmin edebiliyor durumda olmanız beklenmektedir. Buna örnek olarak da 1996’da Deep Model programı Kasparov’u satranç müsabakasında yenmişti. Fakat Kasparov bütün maçları takip edip ezberlemiş ve bunun sonucunda oyun stratejisini belirlemişti. Kasparov alışılmış hamlelerini yapmadığı bir müsabakada program alışık olmadığı hamleleri görünce şaşırmış ve maçı kaybetmiştir.1997 yılına gelindiğindeyse Deep Blue farklı hamleleri ve stratejileri de deneyip Kasparov’u yenmiştir.
2. Production ve Live Performance: Verinin alınması, işlenmesi ve kurduğumuz modele işlenmesini içermektedir. Yie burada hata olması durumunda insan müdahalesi de yapılması gerekecektir.

Ölçeklenebilirlik: Veri çok fazla gerçek mi? Test edildi mi? Hata çıkacak mı? Gibi soruları içermektedir. Domain içinde hangi toolları kullnabilecek düzeyde bilip model oluşturulması ve geliştirme tarafında zaman harcanması önerilmektedir.

Burada yine tez konusunda bazı bilgiler verildi. Kendi kendini süren araçların konvoy oluşturması. Araçların birbiriyle konuşarak şeridi maksimum oranda kullanılması hedeflenmektedir. Karışık trafikte otonom araçları yönlendirerek kullanımlarını maksimize etmeyi planlamaktadır. Gideceği yere göre araçları gruplandırıp yönlendirme ve kontrol kararı vermektedir. Data toplama ise bu işin trajectory yani anlık GPS verisinin sağlanması ve Occupancy tarafında trafik doluluğu ve bu bilgilerin kameralardan alınmasını içermektedir.