

Büyük Veri :

Yapay zekâ teknolojisinin altında büyük veri (big data) yatıyor. Verinin gerçek zamanlı işlenebilmesi, bu datanın farklı kanallardan gelmesi, kendi kendine öğrenen teknolojilere sahip uygun ürünlere ulaşabilmesini sağlamak önemlidir. Artık kural bazlı statik teknolojiler yetmiyor. Yapay zekâyı bu yüzden teknoloji açısından değer yaratacak şekilde kullanıyoruz.

Gelecekte teknoloji departmanlarının yarısından fazlası veri bilimcilerden (Data Scientist) oluşacaktır. Beklenen önemli değişimlerden biri fiziksel dünya ile dijital dünyanın artık iç içe geçmişliği örneğin süt bittiğine otomatik sipariş geçen buzdolaplarından insanların çalışmadığı mağazalara uzanan bir dönüşümden bahsediyoruz. Bu dönüşümün ortasında da yapay zekâ teknolojileri oturuyor.

Şurası çok önemli; çağımızın insanı kendisine özel bir dokunuş istiyor. Yani bir grubun içerisinde addedilerek, o grubun bir parçası olan ortak bir kimlikmiş gibi addedilmek istemiyor. Mutlaka ona özel bir dokunuş” dedim. Tek tek herkese ayrı hitap etmeniz gerekiyor. Mesela ben şahsıma özelleşmemiş hiçbir maili cevaplamam. İsmen hitap edilen, hukukumun gereği olan, karşılığı olan, üslubu olmayan standart hiçbir maili cevaplamam. Ben yokum ki orada. Bir topluluğa atılmış. Kim cevabını veriyorsa versin.

İnsanoğlunun anne karnından itibaren hareket ettiğini ve bunun hayatın sonuna kadar değişmedi. Yapılan bir araştırmaya göre sayfa yüklenme süresindeki her 1 saniyelik hızlanma internet satışlarını yüzde 5 arttırıyor.

Bilgisayar kontrolündeki bir sistemin ya da cihazın faaliyetlerini hafızasındaki verileri kullanarak insan zekasına benzer şekilde yerine getirme yeteneği kazandırılması, veri yığınının öğrenen zekanın geliştirilmesi ile mümkün olabilir. Diğer bir anlatımla, kendi kendine öğrenen matematiksel modeller ve algoritmalar ile insandan bağımsız davranış geliştirmesi gerekmektedir.

Gezgin algılayıcıların, sistemlerin ve makinelerin artması, sosyal ağlara gezgin erişimin yaygınlaşması, çeşitli takip (sensörler, barkodlar, karekodlar, RFID sistemleri... vs.), nesnelerin interneti (IOT) ve otomasyon teknolojilerinin gelişmesi, iletişim teknolojilerinin ulaşılabilirliğinin artması, başta ticari işlemler olmak üzere pek çok iş kolunun elektronik ortama taşınması ile birlikte hem üretilen verinin çeşitliliği hem de toplanma hızı ve miktarı da ciddi oranlarda artmıştır. Bu artış üstel olarak devam etmektedir.

Öte yandan cihazlara takılan sim kartlar, algılayıcılar, elektronik ölçerler, bilgisayar sistemleri ve yazılımları sayesinde, cihazların uzaktan izlenmesini, yönetilmesini ve internet ağı ile birbiriyle iletişim kurabilmesini sağlayan bir teknoloji olan Makineler Arası İletişim (M2M) ve Nesnelerin İnterneti, hem bireylerin hem de şirketlerin hayatında geniş bir kullanım alanı bulmaktadır.

M2M sistemleri, günümüzde neredeyse herhangi bir donanımın çeşitli uygulamalar veya cihazlarla birbirine kolayca bağlanabildiği Nesnelerin İnterneti (IoT), Her şeyin İnterneti (IoE), Nesnelerin Ağı (WoT) ve Her şeyin Ağı (WoE) gibi ortamlara evrilmiştir. Akıllı ortamların meydana geldiği bu sistemlerde, muazzam bir veri hacmi üretilir ve üstelik bu verilerin çoğu yapılandırılmamıştır. Resim, ses, metin, video gibi pek çok türde olabilen ve ağlar üzerinden aktarılan bu veriler bulut ortamlarda da saklanmaya başlamıştır. Bu verilerle ilgili bir başka husus ise sosyal medya verileri gibi insan kaynaklı veriler başta olmak üzere, değişken ve dinamik bir diğer deyişle akan bir yapıya sahip olmalarıdır. Bir yandan sisteme cihazlardan yeni veriler dahil olmakta veya bazı veriler kesintiye uğramakta, öbür yandan mevcut verilerde değişiklik meydana gelebilmektedir. Toplanan verilerin analizi bu sebeple daha karmaşık bir hal almaktadır. “Big data” yani “büyük veri” kavramı bu sebeple de özellikle son yıllarda çokça tartışılır hale gelmiştir.

İnternet ve kablosuz teknolojilerin hayatın her alanında daha çok yer almasıyla birlikte, sayılar, metinler, ifadeler, görüntüler, şekiller, grafikler, sunumlar gibi anlamlı bilgilere dönüşen verilerin devasa miktarlarda depolanması söz konusu olmaktadır. Hareket halindeki mobil uygulamalar ile birlikte, toplanan verinin hem çeşitliliği hem de hacmi inanılmaz boyutlarda artmaya devam etmektedir. Verilerin işlenmesi, sınıflandırılması, analizi, ve anlamlı bilgilere dönüştürülmesi ile kendi kendine karar veren sonuçların elde edilebilmesine de imkân vermektedir. Bu ders çalışmasında, veri analizinde yapay zekâ ve makine öğrenmesi tekniklerinin kullanımı anlatılacaktır. Başlıca yapay zekâ ve makine öğrenmesi teknikleri öğretilecektir. Kümeleme, sınıflandırma, yapay sinir ağları, metin ve web madenciliği, fikir madenciliği, duygu analizi, olasılık ve öğrenerek karar veren algoritmalar konusunda büyük verilerle yapılan çalışmalar anlatılacaktır.

Büyük Veri Analizi

Büyük veri, kısaca 11V diyebileceğimiz beş kavram ile betimlenmektedir:

1) Volume (Hacim),

- 2) Velocity (Hız),
- 3) Variety (Çeşitlilik),
- 4) Verification (Doğrulama)
- 5) Value (Değer).
- 6) Veracity (Gerçeklik),
- 7) Volatility (Oynaklık)
- 8) Validity (Geçerlik)
- 9) Vulnerability (Hassaslık),
- 10) Variability (Değişkenlik),
- 11) Visualization (Görselleştirme).

“Data never sleeps” projesi kapsamında Haziran 2016 verilerine göre, yalnızca bir dakika içerisinde;

- Youtube video paylaşım sitesine kullanıcılar tarafından 400 saatlik video yüklemesinin gerçekleştirildiği,
- Twitter üzerinden 9.678 adet emoji içerikli tweet atıldığı,
- Google’ da 69.500.000 kelime tercüme edildiği, sadece Amerikalı kullanıcıların mobil cihazlarla yaklaşık 18.000 GB veri kullandıkları,
- Facebook Messenger kullanıcılarının 216.302 adet fotoğraf paylaştıkları,
- Instagram kullanıcılarının paylaşılan görüntüler için 2.430.555 adet beğeni yaptıkları
- Amazon web sitesinden 222.283 \$ satış yapıldığı tespit edilmiştir.

Büyük veriden kastedilen yalnızca hacimsel büyüklük değildir. Farklı kaynaklardan ve farklı biçimlerde toplanan verilerin anlamlı ve işlenebilir hale getirilmesi gerekmektedir. Veriler hacim ve tür yanında sürekli artan bir hızda oluşmakta ve depolanmaktadır.

Verilerin depolanma ve değişme hızının yanı sıra büyük verilerin çoğu zaman karmaşık, düzensiz olduğu ve yanlışlar içerebileceği gerçeği, bu verilerin düzenlenmesi ve ayıklanması sorununu doğurmaktadır.

Üstelik anlık alınan verilerden hemen bilginin elde edilmesi yani verinin toplandığı anda analiz edilmesi gerekmektedir. Özellikle internete bağlı cihazların kaynaklık ettiği verilerin analizinde veri madenciliği yöntemleri yanında pazarlama, algı yönetimi, izleme, web, metin ve multimedya madenciliği teknikleri kullanılmaktadır. Burada unutulmaması gereken, verilerin yalnızca dijital değil fotoğraf, resim, video, ses, metin, konum (GPS) bilgisi vs. gibi pek çok çeşitte ve her biri için çeşitli boyutlarda olduğudur. Böyle olunca da asıl önemli olan, bu kadar büyük, hızlı ve çeşitli olan veri topluluğundan anlamlı ve değerli bilgiyi elde etmek olmaktadır.