

Veri Madenciliği Yöntemleri

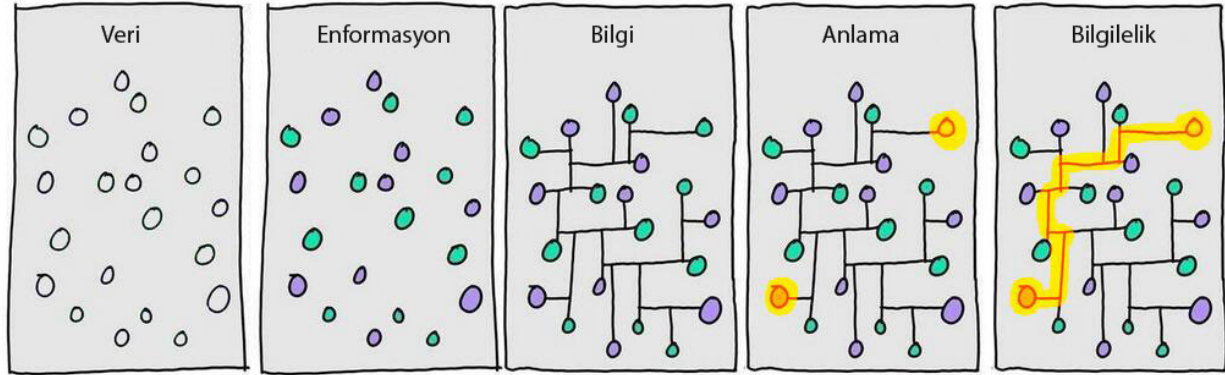
Veri, işlenmemiş ham bilgidir. Örneğin; bir kişinin adı, soyadı, adresi, telefon numarası gibi.

Enformasyon, verinin anlamlı bir konu etrafında işlenmiş halidir. Bir amaca yönelik olarak düzenlenmiştir. Veri anlamlandırıldığı aşamada enformasyon haline gelir.

Örneğin; Bir sınıfın not ortalaması , son 1 yılın en yüksek ortalamasıdır.

Bilgi , enformasyon haline dönüştürülmüş gerçeklerin analiz edilmesi ve sentezlenmesi sonucu karar vermeye yönelik olarak elde edilen daha üst seviyeli gerçekleri içerir. Bilgi, enformasyonun deney, tecrübe, yorum, analiz ve bağlam ile zenginleştirilmiş halidir.

Verinin kullanılabilir hale getirilmesidir. Örneğin ; Her yıl Zeynep tüm satış temsilcilerinden daha fazla satış gerçekleştirir.



Bilgi hiyerarşisi kendi içinde **veri, enformasyon, bilgi, anlama ve bilgilelik** şeklinde sıralanır.

VERİ İLE BİLGİNİN FARKI

Veriler kullanılarak bilgi oluşturulur. Örneğin; Ali'nin elinde 3 yeşil 2 kırmızı 1 mavi bilye var.

- Veri: 3 yeşil bilye 2 kırmızı bilye 1 mavi bilye
- Bilgi: Ali'de toplam 6 bilye vardır.

Veri arar, Bilgi kullanır. Örneğin ;“Yaşınız nedir? ‘ sorusu bir veridir. '20' cevabı bir bilgidir.

BİLGİ İLE ENFORMASYONUN FARKI

Belirli bir konuda zamanla biriken enformasyon, ayıklanıp sınıflandıktan ve düzenlendikten sonra, genelliği ölçüsünde bilgiye dönüşmektedir. Bu açıdan, enformasyon özel bir konunun anlaşılmasına ya da özel bir problemin çözümüne hizmet ederken, bilgi görece daha genel bir konunun anlaşılması veya belirli türden problemlerin tümünün çözülmesi için kullanılmaktadır.

Örneğin; “Bir ailede 4 kişi var ve araba almaya karar verirler ve arabanın rengi hakkında oylama yaparlar. 1 kişi beyaz 3 kişi siyah der. ” Bu bilgidir. Enformasyon” arabanın siyah alınacağı ” karardır.

ENFORMASYON İLE VERİNİN FARKI

Enformasyon verileri kullanarak anlamlı hale getirir. Enformasyon verileri özetler. *Örneğin ;*

” Bir manav birinci gün kayısıyı 2 liradan satar; ikinci gün 4 liradan satar.” Bu veridir.

“Manav bir günde kayısının fiyatını iki katına çıkarmıştır.” Bu enformasyondur.

Veri Madenciliği Modelleri

Veri madenciliği büyük veri yığınları içerisindeki bilinmeyen bilgiyi ortaya çıkarma işlemidir. Gelecek 5 sene şirketimiz hangi alanda büyüyebilir, bu yıl en iyi 10 müşterimiz kim olacak, hastanın hastalığı ne olabilir gibi.

Veri Madenciliğinde kullanılan modeller, **tahmin edici** (Predictive) ve **tanımlayıcı** (Descriptive) olmak üzere iki ana başlık altında incelenmektedir.

- Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır.
- Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır.

Veri Madenciliği yöntemlerini denetimli ve denetimsiz olmak üzere iki ana kategoriye ayırmak da

mümkündür. Veri Madenciliğinde iyi tanımlanmış veya kesin bir hedef olduğunda denetimli (supervised) ifadesi kullanılır. Elde edilmesi istenen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik söz konusu ise denetimsiz (unsupervised)

ifadesi kullanılır. Denetimli ve denetimsiz ifadeleri birbirinin tersine karşılık gelmektedir. Denetimli ve denetimsiz yöntemleri sürecin bütünü açısından değerlendirmek gerekirse;

- Denetimsiz yöntemler daha çok veriyi anlamaya, tanımaya, keşfetmeye yönelik olarak kullanılan ve sonraki uygulanacak yöntemler için fikir vermeyi amaçlamaktadır,
- Denetimli yöntemler ise veriden bilgi ve sonuç çıkarmaya yönelik kullanılmaktadır, denilebilir. Bu nedenle denetimsiz bir yöntemle elde edilen bir bilgi veya sonucu, eğer mümkünse denetimli bir yöntemle teyit etmek, elde edilen bulguların doğruluğu ve geçerliliği açısından önem taşımaktadır.

