SAKARYA ÜNİVERSİTESİ

Veri Madenciliği Uygulamaları

Hafta 4

Yrd. Doç.Dr. Nilüfer YURTAY

Veri Madenciliğinde Metotlar

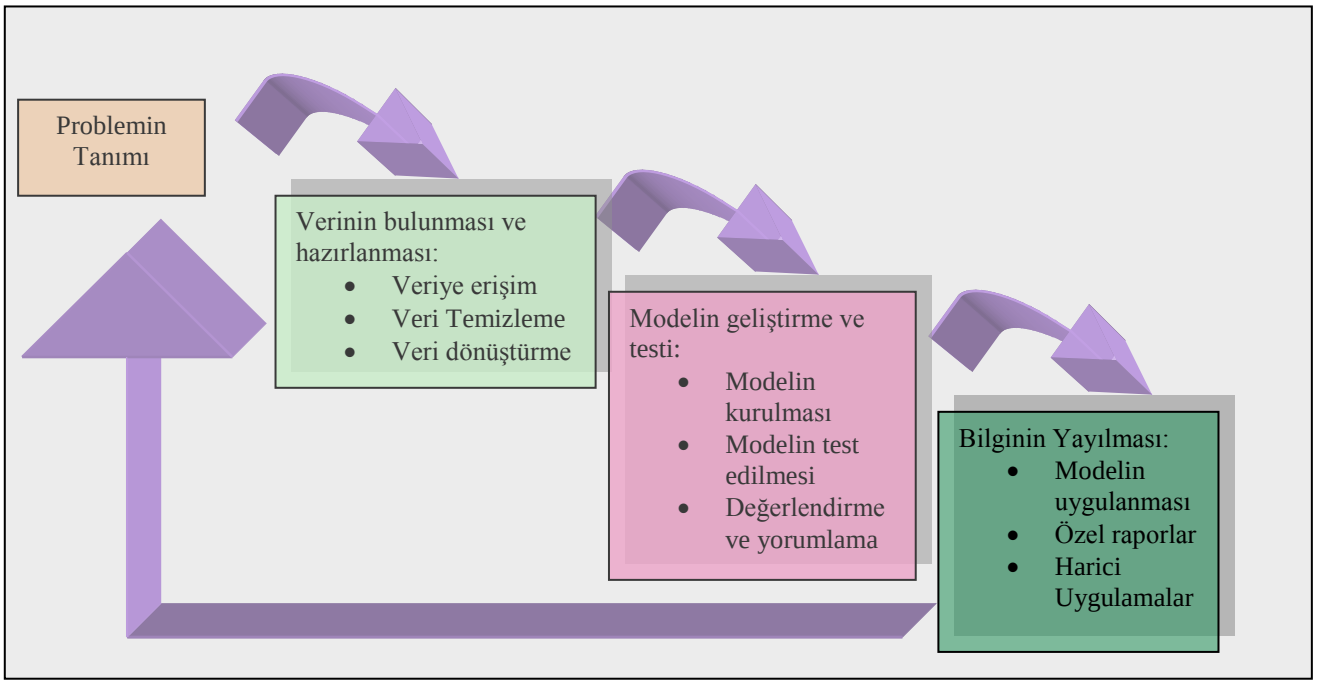
4.1 Giriş

Veri Madenciliğinin Süreci:

- Problemin Tanımı
- Kullanılacak verilerin seçilmesi ve hazırlanması
- Verilerin bulunması ve analizi
- Modelin oluşturulması
- Modelin geçerliliğinin testi
- Bilginin üretilmesi, eylem planına dönüştürülmesi ve sonuçların ölçülüp değerlendirilmesi

olarak değerlendirilebilir(şekil 4.1).

Bir başka deyişle veri madenciliğini verinin temizlenmesi, bütünleştirilmesi, indirgenmesi, dönüştürülmesi, algoritmanın seçimi ve uygulanması , sunum ve değerlendirme biçiminde süreçlerini yazmak da mümkündür



Şekil 4.1 Veri Madenciliği Süreci

4.2 Veri Temizleme

Kayıp, yanlış ve gereksiz verilerin ortadan kaldırılması olarak söylenebilir. Kayıp verilerin ortaya çıkaracağı sorunları ortadan kaldırabilmek için bazı teknikler geliştirilmiştir:

- Kayıp verinin bulunduğu kaydı veri tabanından çıkartmak yada benzer türdeki kayıtları iptal etmek,

- Kayıp verileri elle teker teker doldurmak,
- Tüm kayıp verilere aynı bilgiyi girmek,
- Kayıp olan verilere tüm verilerin ortalama değerinin verilmesi,
- Regresyon yöntemi kullanılarak diğer değişkenlerin yardımı ile kayıp olan verilerin tahmin edilmesi.

4.3 Veri Bütünleştirme

Veri madenciliğinde genellikle farklı veri tabanlarındaki verilerin birleştirilmesi gerekmektedir. Genellikle veri ambarı şeklinde bulunan veriler birleştirilerek üzerinde işlem yapmaya kolaylaştırmaya çalışılır.

Farklı veri tabanındaki veriler birleştirilmesi ile bir takım hataları da beraberinde getirir.

Örneğin, bir veri tabanında girişler "musteri-ID" şeklinde yapılmışken, bir diğerinde "musteri-numarası" şeklinde olabilir.

Bu tip hatalara şema birleştirme hataları denir. Bu tip hatalardan kaçınmak için meta veriler kullanılır.

Veri birleştirmede önemli bir konu da indirgemedir. Bir değişken, başka bir tablodan türetilmişse fazlalık olabilir. Değişkendeki tutarsızlıklar da, sonuçta elde edilen veri kümesinde fazlalıklara neden olabilir. Bu fazlalıklar korelasyon analizi ile araştırılabilir.

Örneğin yukarıda da bahsedilen "musteri-ID" ile "musterinumarası" korelasyon katsayısı bulunabilir.

Eğer bulunan korelasyon katsayısı yüksek bulunuyorsa, değişkenlerden biri veri tabanından çıkarılarak indirgeme yapılır.

4.3.1 Regresyon Denklemleri

İki olayın değişkenleri X ve Y olmak üzere en uygun regresyon denklemini bulmak için X ve Y değerlerinin grafik üzerinde sırasıyla apsis ve ordinat olarak düşünülmesi faydalı olcaktır. Dağılma grafiğindeki bu noktalar belirli bir seyir gösterdikleri takdirde Y=f(X) fonksiyonu regresyon denklemi olacaktır.

Bir başka değişle, iki değişken arasında var olan ilişkinin en uygun bir şekilde hangi matematik fonksiyonla ifade edilebileceğinin aratırılması sonucunda elde edilen denklem regresyon denklemidir. Bu denklem değişkenlerden birinin bilinmesi durumunda diğeri için tahmin yapmanıza imkan verir.

Dağılma grafiğinin noktalarının seyri regresyon denkleminin tipi için önemlidir. Noktalar bir doğru etrafında dağılmış ise iki değişken arasında doğrusal bir ilişki düşünülebilir. Eğer dağılışta bir bükülme noktası varsa bu takdirde iki değişken arasında 2.dereceden bir regresyon denklemi bağlantısı olduğu düşünülebilir. Genelleştirecek olursak;

- Hiç bükülme noktasu yok ise regresyon denklemi doğrusaldır.

$$Y = a + bX$$

- Bir bükülme varsa regresyon denklemi 2.derecedendir.

$$Y = a + bX + cX^2$$

- İki bükülme varsa regresyon denklemi 3.derecedendir.

$$Y = a + bX + cX^2 + dX^3$$

- Üç bükülme varsa regresyon denklemi 4.derecedendir.

$$Y = a + bX + cX^2 + dX^3 + eX^4$$

Yukarıdaki denklemlerin yanısıra aşağıdaki tipde de regresyon denklemleri oluşturulabilir:

$$Y = \frac{1}{a + bX}$$

$$Y = ab^X yada\ logY = loga + Xlogb$$

$$Y = aX^b yada\ logY = loga + blogX$$

4.3.2 Normal Denklemleri

İstatistik olaylar arasındaki ilişkinin en iyi şekilde belirtilebilmesi için, noktalar arasından geçirilecek olan doğru ya da eğrinin bu noktalara olan uzaklıları toplamı minimum olmalıdır. Bu nedenle a,b,c gibi paremetrelerin bu şarta uygun olarak belirlenmesi gerekmektedir.

İstatistiki olaylar arasındaki ilişki doğrusal ise ilgili normal denklemleri şu şekildedir:

$Y = a + bX$ doğru denklemi için

$$\sum Y_i = na + b\sum X_i$$

$$\sum Y_i X_i = a\sum X_i + b\sum X_i^2$$

İstatistiki olaylar arasındaki ilişki 2.dereceden ise ilgili normal denklemleri şu şekildedir:

$Y = a + bX + cX^2$ denklemi için

$$\sum Y_i = na + b\sum X_i + c\sum X_i^2$$

$$\sum Y_i X_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3$$

$$\sum Y_i X_i^2 = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4$$

$X - \bar{X} = x \; ve \; Y - \bar{Y} = y$ olmak üzere

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$ ve $b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$ ifadelerine de sırasıyla y'nin x'e göre ve x'in y'ye göre regresyon katsayıları adı verilir.

X'in belirli bir değerine karşılık gelen Y değerine Y' teorik değeri diyelim. Gerçekleşen değerler ile teorik değerler toplamı birbirine eşit olmakla beraber değerler tek tek karşılaştırıldığında az ya da çok bir fark olacaktır. Bu farklar elde edilmiş olan regresyon doğrusunun gerçek değerlerden mininum uzaklıklarına eşittir. Bu değer yapılan herhangi bir tahminin hatasını gösterir. Formülü şu şekildedir:

$S_{yx} = \sqrt{\frac{\Sigma(Y - Y')^2}{n}}$ bu formül Y'nin X'e göre standart hatasıdır.Benzer şekilde X'in Y'ye göre standart hatası da $S_{xy} = \sqrt{\frac{\Sigma(X - X')^2}{n}}$ biçiminde hesaplanabilir. Burada genellikle $S_{yx} \neq S_{xy}$ dir.

4.3.3. Korelasyon Katsayısı

Herhangi iki olay arasında pozitif ya da negatif bir ilişki söz konusudur. Bu ilişki tam ilişkinin bir yüzdesi olarak belirtilir. İki ya da daha fazla değişkenin arasındaki ilişkinin yönünün ve derecesinin araştırılması korelasyon analizinin konusudur. Korelasyon katsayısı da bu ilişkinin derecesinin tayininde kullanılan bir ölçüdür.

Aşağıdaki formül yardımıyla korelasyon katsayısı hesaplanabilir:

$$r = \frac{\Sigma[(X - \bar{X})(Y - \bar{Y})]}{\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}}$$

Burada $X - \bar{X} = x \; ve \; Y - \bar{Y} = y$ alınırsa

$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$ ya da $r = \frac{\Sigma xy}{n\sigma_x \sigma_{y'}}$ elde edilir.

Korelasyon katsayısının değeri -1 ile +1 arasındadır. +1 olduğunda değişkenler arasında ikişki tam ve pozitifdir denir. -1 için de tersi söz konusudur. Korelasyon değerinin ∓ 1 e yakın olması ilişkinin kuvvetli olduğuna, 0'a yakın olması ise ilişkinin zayıf olduğuna işaret eder.

Korelasyon katsayısı regresyon katsayılarının aritmetik ortalamasıdır.

$$r = \mp\sqrt{b_{xy} b_{yx}}$$

Regresyon katsayıları pozitif ise korelasyon katsayısı da pozitif, Regresyon katsayıları negatif ise korelasyon katsayısı da negatifdir.

Son olarak tahminin standart hatasının formülle hesaplanışında kullanılan formülü verelim:

$S_{yx} = \sqrt{\frac{\Sigma y^2 - b_{yx}\Sigma xy}{n}}$ ve $S_{xy} = \sqrt{\frac{\Sigma x^2 - b_{xy}\Sigma xy}{n}}$ biçimindedir.

Örneğin aşağıdaki tabloda 5 ailenin çocuk sayıları ve anne yaşları gösterilmektedir:

Aile No	Çocuk sayıları(X)	Anne yaşları(Y)
1	2	25
2	1	20
3	5	35
4	4	45
5	3	25
TOPLAM	15	150

a) Y'nin X'e göre regresyon doğrusunu bulalım.

$Y = a + bX$ denkleminde a ve b katsayılarını bulmak yeterlidir.

$\sum Y_i = na + b\sum X_i \quad \sum Y_i X_i = a\sum X_i + b\sum X_i^2$ normal denklemlerini kuralım.

Aile No	Çocuk sayıları(X)	Anne yaşları(Y)	X^2	XY
1	2	25	4	50
2	1	20	1	20
3	5	35	25	175
4	4	45	16	180
5	3	25	9	75
TOPLAM	15	150	55	500

$$\sum Y_i = na + b\sum X_i \Rightarrow 150 = 5a + 15b$$

$$\sum Y_i X_i = a\sum X_i + b\sum X_i^2 \Rightarrow 500 = 15a + 55b$$

Denklemlerin ortak çözülmesiyle a=15 ve b=5 bulunur. Buna göre regresyon doğru denklemi $Y = 15 + 5X$ olur.

b)Anne yaşları ile ilgili teorik değerleri elde edilen regresyon doğrusu yardımıyla tespit edelim.

$X = 2$ için $Y' = 15 + 5.2 = 25$
$X = 1$ için $Y' = 15 + 5.1 = 20$
$X = 5$ için $Y' = 15 + 5.5 = 40$

$$X = 4 \text{ için } Y' = 15 + 5.4 = 35$$
$$X = 3 \text{ için } Y' = 15 + 5.3 = 30$$

c) Teorik değerlere dayanarak tahminin standart hatasını bulalım.

Anne yaşları(Y)	Y'	Y-Y'	$(Y-Y')^2$
25	25	0	0
20	20	0	0
35	40	-5	25
45	35	10	100
25	30	-5	25
150	150	0	150

$$S_{yx} = \sqrt{\frac{\Sigma(Y - Y')^2}{n}} = \sqrt{\frac{150}{5}} = 5.47$$

d) Tahminin standart hatasını formül yardımıyla bulalım.

$$S_{yx} = \sqrt{\frac{\Sigma y^2 - b_{yx}\Sigma xy}{n}}$$

Çocuk sayıları(X)	Anne yaşları(Y)	X-Xort=x	Y-Yort=y	xy	y^2
2	25	-1	-5	5	25
1	20	-2	-10	20	100
5	35	2	5	10	25
4	45	1	15	15	225
3	25	0	-5	0	25
15	150	0	0	50	400
Xort=3 ve Yort=30 dir.					

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{50}{10} = 5, \qquad S_{yx} = \sqrt{\frac{400 - 5.50}{5}} = 5.47$$

e) Son olarak r korelasyon katsayısını bulalım.

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{50}{10} = 5 \quad \text{ve} \quad b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{50}{400} = 0{,}125$$
$$r = \mp\sqrt{b_{xy}b_{yx}} = \mp\sqrt{0.125 * 5} = 0.79$$

4.4 Veri İndirgeme

Veri indirgeme teknikleri, daha küçük hacimli olarak ve veri kümesinin indirgenmiş bir örneğinin elde edilmesi amacıyla uygulanır. Bu sayede elde edilen indirgenmiş veri kümesine veri madenciliği teknikleri uygulanarak daha etkin sonuçlar elde edilebilir. Temel olarak boyut indirgeme ve satır indirgeme şeklinde iki yönde gerçekleştirilir.

Veri indirgeme yöntemleri aşağıdaki biçimde özetlenebilir :

1. Veri Birleştirme veya Veri Küpü (Data Aggregation or Data Cube)
2. Veri Sıkıştırma (Data Compression)
3. Kesikli hale getirme (Discretization)
4. Boyut indirgeme (Dimension Reduction)

Veri birleştirme veya veri küpü yapılacak 2000-2003 yılları için çeyrek dönemlik satış tutarlarından oluşan bir veri kümesinin bulunduğunu varsayalım. Bu yıllar için yıllık satış tutarları tek bir tabloda toplandığında veri birleştirmesi yapılmış olur. Sonuç olarak elde edilen veri kümesinin hacmi daha küçüktür fakat yapılacak analiz için bir bilgi kaybı söz konusu değildir. Veri küpleri ise çok değişkenli birleştirilmiş bilginin saklandığı küplerdir. Örneğin bir firmanın satış tutarları yıllar, satışı yapılan ürünler ve firmanın farklı satış yerleri için aynı küp üzerinde gösterilebilir. Veri küpleri özet bilgiye herhangi bir hesaplama yapmadan hızlı bir biçimde erişilmesini sağlarlar.

Veri sıkıştırmada ise orijinal verileri temsil edebilecek indirgenmiş veya sıkıştırılmış veriler, veri şifreleme veya dönüşümü ile elde edilirler. Bu şekilde indirgenmiş veri kümesi, orijinal veri kümesini bir bilgi kaybı olacak biçimde temsil edebilecektir. Bununla beraber bilgi kaybı olmaksızın indirgenmiş veri kümesi elde edilmesine yarayacak bir takım algoritmalar da mevcuttur. Bu algoritmalar bir takım sınırlamalara sahip olduklarından sıkça kullanılamamaktadır. Bununla beraber temel bileşenler analizi gibi yöntemler, bir bilgi kaybına göz yumularak sıkıştırılmış veri kümesi elde edilmesinde kullanışlıdır.

Analiz edilecek veri miktarını azaltmak için veri sıkıştırılır. Eğer veri sıkıştırma sonrası orijinal veriden herhangi bir bilgi kaybı oluşmuyorsa buna kayıpsız (**lossless**) sıkıştırma adı verilir. Ancak orijinal veriye yaklaşık bir değer elde (belli bir oranda kayıp varsa) ediliyorsa buna da **lossy** sıkıştırma denir. Sıkıştırma işlemleri belirli algoritmalara göre yapılır. Sıklıkla kullanılan sıkıştırma yöntemleri: *Wavelet Transforms* ve *Temel Bileşen Analizi*.

Wavelet Transforms (DWT – discrete wavelet transforms)
Veriyi, orijinal veriyle aynı uzunlukta farklı bir vektöre çevirir. Bu çevrilen yeni vektör de kısaltılabilir – kesilebilir olduğundan veri, işimize yarayacak kadar azaltılmış olur. Ayrıca smoothing (düzeltme) işlemi yapmadan verinin gereksiz bilgilerden temizlenmesi görevini de yerine getirir. DWT küp yapıdaki çok boyutlu verilere uygulanabilir. DWT nin ne kadar kompleks olduğu küpteki hücre sayısına bağlıdır. DWT ayrık ve çarpık verilerle daha iyi sonuçlar verir ve gerçek hayatta da sıklıkla kullanılmaktadır. (parmakizi resimlerinin sıkıştırılması, zaman serisi veri analizi, veri temizleme)

Temel Bileşen Analizi (PCA - principal component analysis)
Verinin yapısını bozmadan veriden belirli sayıdaki kaydı ve belli boyuta göre alır ve analiz eder. Hem kayıt sayısında azalma olur, hem de kolon – boyut indirgenir. PCA da veri azaltmak için orijinal veriden n tane kayıt k tane boyut ortagonal olarak bulunur. İşleyişinde veri normalize edilir. PCA orijinal veriyle aynı yapında n tane ortagonal vektör hesaplar. Bu vektörlere temel bileşen denir. Vektörler sıralanır. Sıralanan bu vektörlere göre en zayıf bileşen çıkartılarak data azaltılır. PCA nın uygulanması ucuzdur. Yüksek performans beklenmez. Sıralı – sırasız kolonlara, veri özelliklerine uygulanabilir. Boyut sayısı arttığında DWT yi kullanmak daha çok tercih edilebilir.

Kesikleştirme ise bazı veri madenciliği algoritmaları yalnızca kategorik değerleri ele aldığından, sürekli verilerin kesikli değerlere dönüştürülmesini içerir. Bu şekilde sürekli verilerin kesikli değer aralıklarına dönüştürülmesiyle elde edilen kategorik değerler, orijinal veri değerlerinin yerine kullanılırlar. Bir kavram hiyerarşisi (concept hierarchy), verilen sürekli değişken için, değişkenin ayrıştırılması olarak tanımlanabilir. Kavram hiyerarşileri, düşük düzeyli kavramların yüksek düzeyli kavramlarla değiştirilmesiyle verilerin indirgenmesinde kullanılır. Örneğin yaş değişkeni 1-15, 16-40,

40+ olacak biçimde daha yüksek kavram düzeyinde ifade edilebilir. Bu şekilde veri indirgemede detay bilgiler kayboluyorsa da, genelleştirilmiş veriler daha anlamlı olacak, daha kolay yorumlanabilecek ve orijinal verilerden daha düşük hacim kaplayacaktır. Kullanılan veri madenciliği programları sayılan veri ön işleme tekniklerinden bir çoğunu gerçekleştirmektedir. Bununla beraber veri işleme ile ilgili özel programlar veya veri ön işleme açısından güçlü bir takım özel programlar vardır. Özellikle veri ön işleme tekniklerini içeren bu açıdan güçlü programlar arasında; BioComp i-Suite, Data Digest Business Navigator 5, Data Detective, IBM Intelligent Miner for Data, KXEN, Magnify PATTERN, Quadstone DecisionHouse, Salford Systems Data Mining Suite ve Xpertrule Miner 4.0 sayılabilir.

Veri madenciliği yapılacak veri kümesi bazen gereksiz olarak yüzlerce değişken içerebilir. **Örneğin** bir ürünün satışına ilişkin olarak düzenlenen bir veri kümesinde, tüketicilerin telefon numaraları gereksiz bir değişken olarak yer alabilir. Bu tür gereksiz değişkenler elde edilecek örüntüleri kalitesizleştirebileceği gibi veri madenciliği sürecinin yavaşlamasına da yol açacaktır. Gereksiz değişkenlerin elenmesi amacıyla ileri veya geri yönlü olarak sezgisel seçimler yapılabilir. İleri yönlü sezgisel seçimde orijinal değişkenleri en iyi temsil edecek değişkenler belirlenir. Ardından her bir değişken veya değişkenler grubunun, bu kümeye dahil edilip edilmeyeceği sezgisel olarak belirlenir. Geri yönlü sezgisel seçimde ise öncelikle değişkenlerin tüm kümesi ele alınır. Daha sonra gereksiz bulunan değişkenler kümeden dışlanarak, en iyi değişken kümesi elde edilmeye çalışılır.

Boyut indirgeme amacıyla kullanılacak bir diğer yöntem ise karar ağaçlarıdır. Karar ağaçları ele alınacak çıktı değişkenini en iyi temsil edecek değişken kümesini verecektir. Veri sıkıştırmada ise orijinal verileri temsil edebilecek indirgenmiş veya sıkıştırılmış veriler, veri şifreleme veya dönüşümü ile elde edilirler. Bu şekilde indirgenmiş veri kümesi, orijinal veri kümesini bir bilgi kaybı olacak biçimde temsil edebilecektir. Bununla beraber bilgi kaybı olmaksızın indirgenmiş veri kümesi elde edilmesine yarayacak bir takım algoritmalar da mevcuttur. Bu algoritmalar bir takım sınırlamalara sahip olduklarından sıkça kullanılamamaktadır. Bununla beraber temel bileşenler analizi gibi yöntemler, bir bilgi kaybına göz yumularak sıkıştırılmış veri kümesi elde edilmesinde kullanışlıdır.

Kolon – boyut indirgenmesinde uygulanacak 3 yol vardır:

1-İleri doğru seçim: Boyutlar taranarak belli istatistiki bilgiler işe yarayacağı düşünülerek seçilir.
2-Tersten eleme: Boyutlara geriden bakarak işe yaramayan bilgiler çıkarılır.
3-İleri doğru seçim ve tersten eleme kombinasyonu: İstatistiki bilgilerden en iyilerini seçerek, en kötülerini eleyen yöntemdir

Boyut indirmedeki amaç boyut fazlalığını boşa çıkartarak veri madenciliği bakımından ihtiyac duyulan bellek ve zaman miktarını azaltmaktır. Boyut indirgeyerek daha kolay gösterimler sağlanabilir. (örn. çok boyutlu uzay üç boyuta düsürülerek görsellestirme araçları ile veriler görsellestirilebilir) . Ayrıca iliskisiz özellikleri elemeye veya gürültüyü azaltmaya yardımcı olabilir. (belli bir eşiğin altında kalan olasılığa sahip veriler dikkate alınmama gibi.)

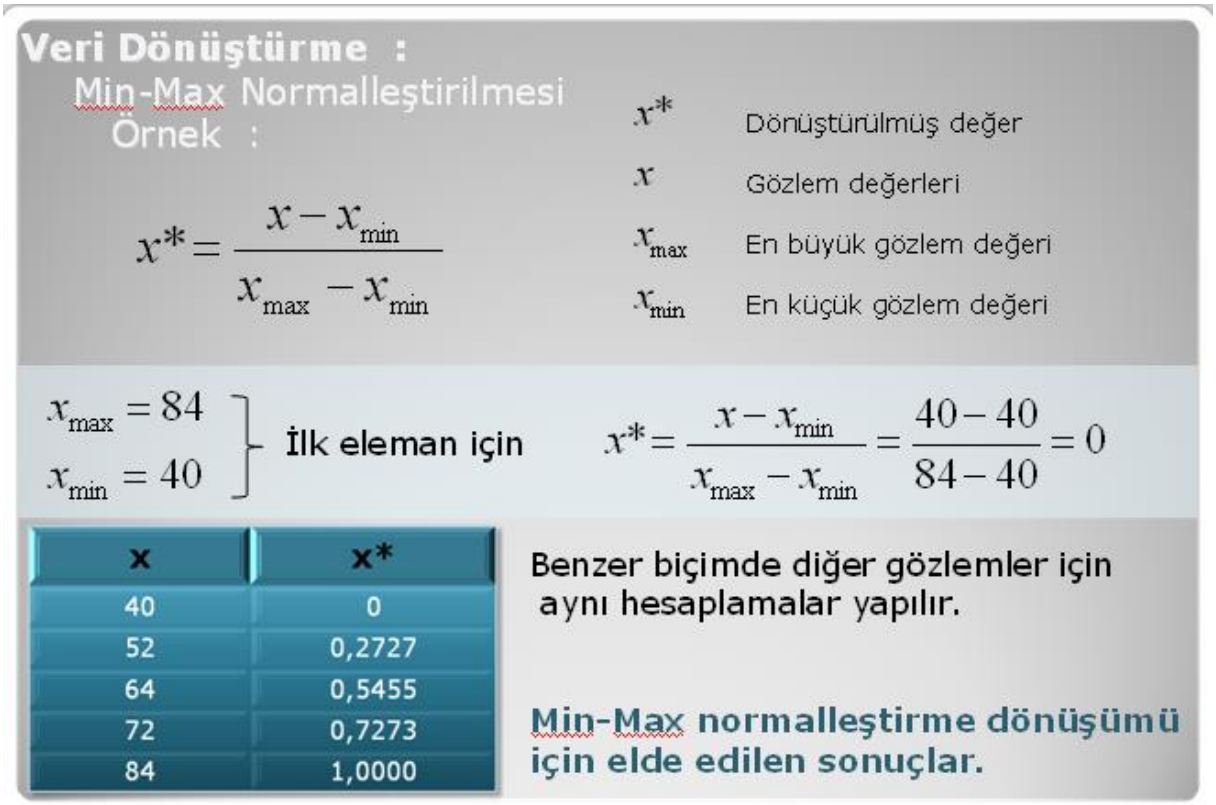
4.5 Veri Dönüştürme

Veri dönüştürme ile veriler, veri madenciliği için uygun formlara dönüştürülürler. Veri dönüştürme; düzeltme, birleştirme, genelleştirme ve normalleştirme gibi değişik işlemlerden biri veya bir kaçını içerebilir. Veri normalleştirme en sık kullanılan veri dönüştürme işlemlerinden birisidir.

Veri normalleştirme tekniklerinden bazıları aşağıdaki biçimde sıralanabilir (Roiger and Geatz, 2003:156):

1. Min-Max
2. Z Skor
3. Ondalık Ölçekleme

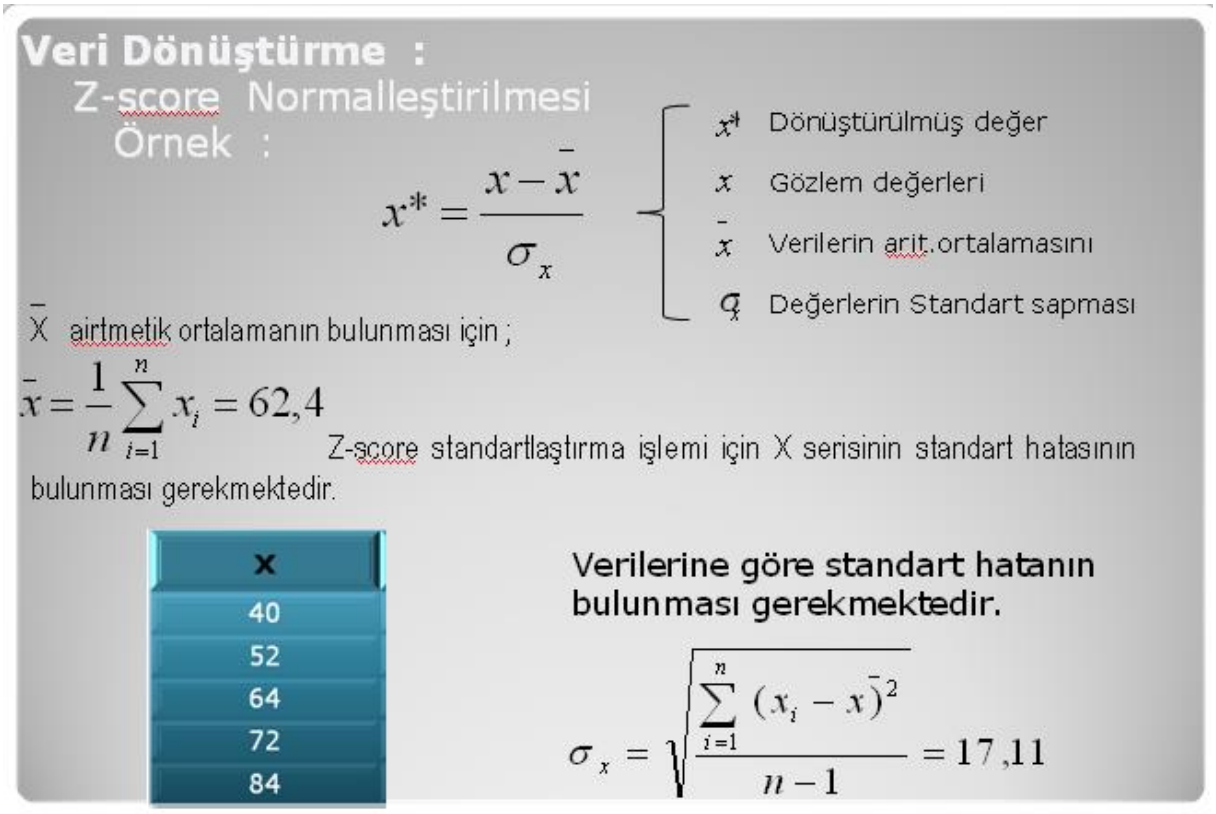
Min-max normalleştirmesi ile orijinal veriler yeni veri aralığına doğrusal dönüşüm ile dönüştürülürler. Bu veri aralığı genellikle 0-1 aralığıdır(Şekil 4.2).



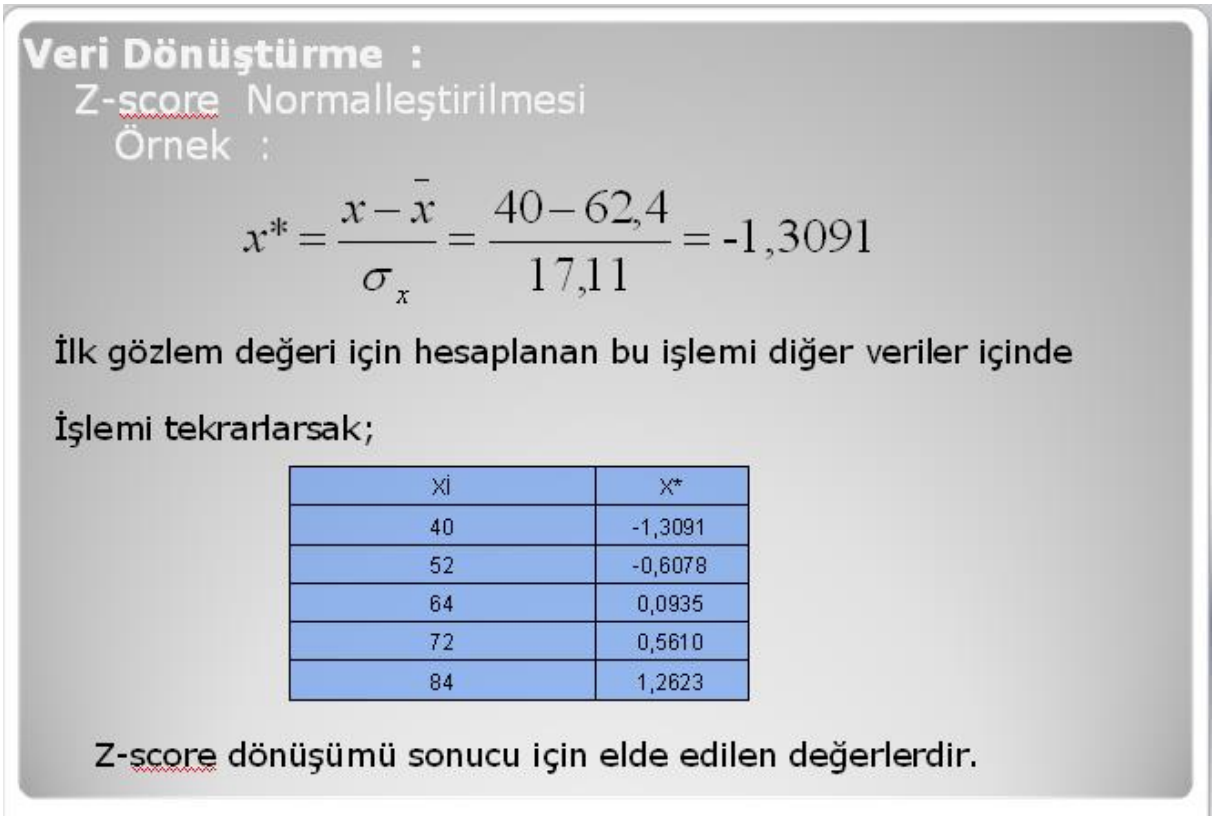
Şekil 4.2 Min-Mak Normalleştirmesi

Z Skor normalleştirmede (veya 0 ortalama normalleştirme) ise değişkenin her hangi bir y değeri, değişkenin ortalaması ve standart sapmasına bağlı olarak bilinen Z dönüşümü ile normalleştirilir (Şekil 4.3-Şekil 4.4).

Ondalık ölçekleme ile normalleştirmede ise, ele alınan değişkenin değerlerinin ondalık kısmı hareket ettirilerek normalleştirme gerçekleştirilir. Hareket edecek ondalık nokta sayısı, değişkenin maksimum mutlak değerine bağlıdır. Ondalık ölçeklemenin formülü aşağıdaki şekildedir: Örneğin 900 maksimum değer ise, n=3 olacağından 900 sayısı 0,9 olarak normalleştirilir.



Şekil 4.3 Z-score Normalleştirmesi



Şekil 4.4 Z-score Normalleştirmesi(devam)

4.6 Algoritma Seçimi

Veri madenciliği yöntemlerini uygulayabilmek için veriler üzerinde yapılan işlemlerin gerekli olanları uygulanır.

Veri hazır hale getirildikten sonra konuyla ilgili veri madenciliği algoritmaları uygulanır.
Söz konusu kullanılan yöntemler ve algoritmalar :

Kullanılan Yöntemler :

- Sınıflandırma
- Kümeleme
- Görselleştirme
- İlişki kurma
- Tahmin modelleri

Kullanılan Algoritmalar :

- Sinir Ağları (neural networks)
- Karar Ağaçları (decision trees)
- Genetik Algoritmalar
- İstatistiksel Analiz

4.7 Sunum ve Değerlendirme

Bilginin görsel bir formata eşleşmesi mümkün müdür? Sorusuna evet cevap verilemesi beklenir. Veri nesneleri, onların öznitelikleri ve veri nesneleri arasındaki ilişkiler; noktalar, satırlar,şekiller ve renkler gibi grafiksel elemanlara dönüştürülebilir. Örneğin

– Nesneler sıklıkla noktalarla sunulur.
– Onların özellikleri noktaların karakteristikleri (renk,boyut ve şekil gibi) veya pozisyonu olarak sunulabilir.
– Eğer pozisyon bilgisi kullanılırsa taşmalar ve grup içinde kalmalar rahatça izlenebilir ve kolaylıkla taşma tespiti algılanabilir.

Kaynaklar

- Veri Madenciliği DR Gökhan Silahtaroğlu 06’2008
- Veri Madencilği Yöntemleri Dr. Yalçın Özkan 06’2008
- M.A.Duchaineau, M.Wolinsky, D.E.Sigeti, M.C. Miller, C. Aldrich and M.B.Mineev - Weinstein,
- “ROAMingTerrain: Real-time Optimally Adapting Meshes”. IEEE Visualization’97, 81–88. Nov. 1997
- Kitap : Introduction to Data Mining, Pang-Ning Tan, Michigan State University, Michael Steinbach, University of Minnesota, Vipin Kumar, University of Minnesota