

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**VERİ MADENCİLİĞİ İLE METEOROLOJİK
PARAMETRELERİN ANALİZİ ve ZİRAÎ METEOROLOJİ
HARİTASININ ÇIKARILMASI**

Fatih TOPALOĞLU

Tez Yöneticisi
Yrd.Doç.Dr. Ahmet ÇINAR

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ELAZIĞ, 2007

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**VERİ MADENCİLİĞİ İLE METEOROLOJİK
PARAMETRELERİN ANALİZİ ve ZİRAÎ METEOROLOJİ
HARİTASININ ÇIKARILMASI**

Fatih TOPALOĞLU

Yüksek Lisans Tezi
Bilgisayar Mühendisliği Anabilim Dalı

Bu tez, 15 / 08 / 2007 tarihinde aşağıda belirtilen jüri tarafından oybirliği /oyçokluğu ile başarılı / başarısız olarak değerlendirilmiştir.

Danışman: Yrd.Doç.Dr. Ahmet ÇINAR

Üye: Yrd.Doç.Dr. Arif GÜLTEN

Üye: Yrd.Doç.Dr. Ali KARCI

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../..... tarih ve sayılı kararıyla onaylanmıştır.

TEŞEKKÜR

Bu tez çalışması boyunca ilgi ve yardımlarını esirgemeyen danışman hocam, Sayın Yrd. Doç. Dr. Ahmet ÇINAR ve Sayın Yrd.Doç.Dr. A.Bedri ÖZER hocalarıma teşekkür ve şükranlarımı sunarım.

İÇİNDEKİLER

İÇİNDEKİLER	I
ŞEKİLLER LİSTESİ	III
TABLolar LİSTESİ	IV
ÖZET	V
ABSTRACT	VI
1. GİRİŞ	1
1.1 Tezin Amacı	1
1.2 Tezin İçeriği	2
2. VERİ MADENCİLİĞİ	4
2.1 Veri Ambarının Oluşturulması	4
2.2 Veri Madenciliği Modelleri	5
2.3 Veri Madenciliğinde Kullanılan Algoritmalar	6
2.3.1 İlişki Bulma Algoritması	6
2.3.2 Sınıflandırma Algoritması	10
2.3.2.1 Karar Ağaçları ile Sınıflandırma	11
2.3.2.2 Yapay Sinir Ağları ile Sınıflandırma	11
2.3.2.3 Bayes Teoremi ile Sınıflandırma	13
2.3.2.4 Bayes Ağları ile Sınıflandırma	14
2.3.2.5 Modellerin değerlendirilmesi	15
2.3.3 Kümeleştirme Algoritması	16
2.3.3.1 Bölünmeli Yöntemler ile Kümeleme	16
2.3.3.2 Hiyerarşik Yöntemler ile Kümeleme	19
2.3.3.3 Yoğunluk Tabanlı Yöntemler ile Kümeleme	21
2.3.3.4 Model Tabanlı Yöntemler ile Kümeleme	22
2.3.4 Ardışık Desenler Algoritması	23
2.3.5 Benzer Seriler Algoritması	23
2.3.6 Tahmin Algoritması	24
3. VERİ MADENCİLİĞİNİN SÜRECİ	25
3.1. Problemin Tanımlanması	26
3.2. Verilerin Hazırlanması	26
3.2.1 Toplama	26
3.2.2 Birleştirme ve Temizleme	26
3.2.3 Dönüştürme	27
3.3. Modelin Kurulması ve Değerlendirilmesi	27
3.4. Modelin Kullanılması	27

3.5. Modelin İzlenmesi.....	27
4. MALATYANIN ZİRAİ METEOROLOJİ HARİTASININ ÇIKARILMASI.....	28
4.1. Problemin Tanımlanması	28
4.2. Verilerin Hazırlanması	28
4.2.1 Toplama.....	29
4.2.2 Birleştirme ve Temizleme	36
4.2.3 Dönüştürme	37
4.3. Modelin Kurulması ve Değerlendirilmesi.....	37
4.4. Modelin Kullanılması.....	42
4.5. Modelin İzlenmesi.....	51
5.SONUÇ	52
5.1. Sonuçlar.....	52
5.2. Öneriler	52
KAYNAKLAR	53
ÖZGEÇMİŞ	54

ŞEKİLLER LİSTESİ

Şekil 2.1	Veri madenciliği modelleri	5
Şekil 2.2	Yaygın öğeleri belirleme	8
Şekil 2.3	İlişkilendirme kuralları oluşturma	8
Şekil 2.4	Yaygın öğe adaylarının bulunması	9
Şekil 2.5	Yaygın öğe azaltma yöntemi	10
Şekil 2.6	Karar ağaç yapısı	11
Şekil 2.7	Yapay sinir ağı uygulaması.....	12
Şekil 2.8	Yapay sinir ağı.....	13
Şekil 2.9	YSA oluşturma	14
Şekil 2.10	YSA ağırlıklarını öğrenme.....	14
Şekil 2.11	Bayes ağları	15
Şekil 2.12	Veri kümesi.....	17
Şekil 2.13	Bölünmeli demetleme	17
Şekil 2.14	K-Metod uygulaması	18
Şekil 2.15	Merkez nokta seçimi	18
Şekil 2.16	Hiyerarşik yöntem.....	19
Şekil 2.17	Aşağıdan yukarıya doğru kümeleme (AGNES)	20
Şekil 2.18	Yukardan aşağı doğru kümeleme (DIANA)	20
Şekil 2.19	DBSCAN	21
Şekil 2.20	Erişebilir nesne	22
Şekil 2.21	Yoğunluk bağlantılı nesne	22
Şekil 3.1	Veri madenciliği standart süreci	25
Şekil 4.1	Hiyerarşik kümeleme yöntemi.....	38
Şekil 4.2	AGNES kümeleme yöntemi	38
Şekil 4.3	Tarım ürünleri kümeleme(AGNES)	39
Şekil 4.4	Karar ağaç yapısı	40
Şekil 4.5	Tarım ürünleri karar ağaç yapısı	40

TABLolar LİSTESİ

Tablo 2.1 Market sepet verisi	6
Tablo 2.2 Yapay sinir ağı uygulaması	12
Tablo 4.1 Malatya ili ortalama sıcaklık değerleri	29
Tablo 4.2 Malatya ili ortalama maksimum sıcaklık değerleri	30
Tablo 4.3 Malatya ili ortalama minimum sıcaklık değerleri	31
Tablo 4.4 Malatya ili nisbi nem değerleri	32
Tablo 4.5 Malatya ili güneşlenme süresi değerleri	33
Tablo 4.6 Malatya ili yağış değerleri	34
Tablo 4.7 Malatya ili rüzgar hızı değerleri	35
Tablo 4.8 Tarım ürünleri iklim istekleri değerleri	36
Tablo 4.9 Ürünlerin yetişmesini sağlayan etkin ay aralıkları	36
Tablo 4.10 Meteorolojik verilerin dönüştürme aşamasındaki gösterimi	37
Tablo 4.11 Tarım ürünlerinin 9 kümede gösterimi	39
Tablo 4.12 Tarım ürünlerinin sınıflandırılması	41
Tablo 4.13 Erzurum ili meteorolojik parametreleri	42
Tablo 4.14 Erzurum ili tarımsal ürün sınıflandırılması	43
Tablo 4.15 Iğdır ili meteorolojik parametreleri	43
Tablo 4.16 Iğdır ili tarımsal ürün sınıflandırılması	44
Tablo 4.17 Apriori Algoritmasında Kullanılan Değişkenler	44
Tablo 4.18 Apriori Algoritması Kesiti	45
Tablo 4.19 Apriori-gen Aday Küme Üretme Algoritma Kesiti	45
Tablo 4.20 Tarım ürünleri verisi	46
Tablo 4.21 İtemlerin support değerleri	47
Tablo 4.22 Tekli birlikteliklerin min.support değerine göre düzenlenmesi	48
Tablo 4.23 İkili birlikteliklerin oluşturulması	48
Tablo 4.24 İkili birlikteliklerin min.support değeri dikkate alınarak düzenlenmesi	49
Tablo 4.25 Üçlü birlikteliklerin oluşturulması	49
Tablo 4.26 Buğday, Çavdar ve Pancar için birliktelik kurallarının alt kümesi	50
Tablo 4.27 Doğu anadolu Bölgesi tarımsal ürün desen keşfi	51

ÖZET

Yüksek Lisans Tezi

VERİ MADENCİLİĞİ İLE METEOROLOJİK PARAMETRELERİN ANALİZİ ve ZİRAÎ METEOROLOJİ HARİTASININ ÇIKARILMASI

Fatih TOPALOĞLU

Fırat Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

2007, Sayfa: 54

Veri madenciliği, büyük veri kaynaklarından önceden bilinmeyen gizli, önemli ve yararlı bilgileri çıkartmaktır. Bu çalışmada, Malatya'nın ziraî meteoroloji haritasının çıkartılması suretiyle tarımsal ürünlerin desen keşfinin yapılması amaçlanmıştır. Ayrıca üreticinin yapacağı yatırımı önceden planlayabilmesi ve bu şekilde yanlış yatırımlardan doğacak maddi kayıpları en aza indirmesi ve yapılacak büyük projelerde ülke ekonomisine katkıda bulunulması hedeflenmiştir.

Veri madenciliği teknikleri kullanılarak, büyük miktarda verinin saklandığı meteoroloji ve tarımsal veri kaynakları veri madenciliği modellerinden kümeleme, sınıflandırma ve ilişkilendirme modelleriyle işlenip yorumlanması sağlanmıştır. Bunun sonucunda Malatya'nın ziraî meteoroloji haritası çıkartılması ve tarımsal ürünlerin desenlerinin keşfinin bulunması sağlanmıştır. Kümeleme modeli ile ürünlerin yapısal özelliklerine göre hangi kümeye ait olduğu, sınıflandırma modeli ile ürünlerin verim analizlerine göre önceden belirlenmiş sınıflardan hangisi ile temsil edileceği, ilişkilendirme modeli ile de ürünler arasındaki gizli ilişkilerin (desenlerin) ortaya çıkarılması sağlanmıştır.

Uygulama olarak , Malatya'nın ziraî meteoroloji haritası çıkartılmıştır. Ayrıca girilen bir ürünün hangi kümeyi ait olduğu, bu küme içerisinde hangi sınıfla temsil edileceği ve bu ürünün varlığıyla diğer yetiyecek ürünlerin varlıklarının tahmin edilmesi sağlanmıştır.

Anahtar Kelimeler: Veri Madenciliği, Karar Ağaçları, Hiyerarşik Yöntem, Apriori Algoritması

ABSTRACT

Master Thesis

ANALYSIS OF METEOROLOGICAL PARAMETERS AND PREPARATION OF AGRICULTURAL METEOROLOGY MAPS BY USING DATA MINING

Fatih TOPALOĞLU

Fırat University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

2007, Page: 54

Data mining is to infer important and useful data which are unknown before from great data sources. In this study, the aim is to discover the pattern of agricultural products by means of drawing agricultural meteorology map of Malatya. Besides, it is intended that producers can plan for their investments in advance, thus, minimize the financial loss to be caused by misstep, and finally contribute to the economy with great projections.

Using data mining techniques, meteorological and agricultural data sources in which a great deal of data have been stored can be processed and interpreted with Classification, Relating and Grouping, all of which are data mining models. In the end, agricultural meteorology map of Malatya has been drawn and pattern of agricultural products has been discovered. Grouping model reveals which group products belong to according to their features. Classification model reveals with which predefined classes products will be presented according to productivity analysis. Relating model reveals hidden relationships among products.

In practice, agricultural meteorology map of Malatya has been drawn. It is achieved which group the product that is entered belong to and with which class the product that is entered will be presented. The other products which will grow dependently upon the product that is entered are estimated.

Keywords : Data Mining, Decision trees, Hierarchy Method, Apriori Algorithm

1.GİRİŞ

Karar aşamalarında çok kritik bazı bilgiler vardır ki, sonuçların etkileri bu bilgilerin doğruluğuyla orantılıdır. Veri madenciliği, organizasyonların karar aşamaları için veri içerisinden önceden bilinmeyen yararlı bilgileri dinamik bir süreçle çıkartıp gelecekle ilgili tahminler ve planlar yapmamızı sağlayan bir dizi teknikler ve anlayışlar bütünü olarak tanımlanabilir.[1]

Veri madenciliği bir çok alanda olduğu gibi meteoroloji ve atmosfer bilimleri alanında da sıklıkla kullanılmaktadır. Meteoroloji ve atmosfer bilimleri alanında bölgesel iklim, yağış haritaları oluşturma, zirai meteoroloji haritaları çıkarma, hava tahminleri, ozon tabakasındaki deliklerin tespiti, çeşitli okyanus hareketlerinin belirlenmesi, rüzgâr ve güneş enerjisi tribünlerinin kurulumunun belirlenmesinde kullanılmaktadır.

Günümüzde artık neredeyse tüm meteoroloji cihazlarının dijital hale gelmesi bu alanda aşırı veri birikmesi sonucunu doğal hale getirmiştir. Kağıt üzerinde veri toplanan klasik meteoroloji bilgi sistemlerinden farklı olarak buradaki verilerden yararlanmak her ne kadar çok daha kolay gibi görünse de aslında diğer alanlardaki veriler gibi bunlarında bireysel çalışmalarla işlenmesi ve yorumlanması neredeyse imkansız denecek kadar zordur.[2]

Bu aşamada sıkıntıları gidermek için veri madenciliği teknikleri yetişmektedir. Bu sayede aşırı miktardaki verinin işlenip yorumlanması mümkün hale gelmiştir. Sadece bunun mümkün olması değil belki de klasik yöntemlerle bulunması çok zor veya imkansız olan bazı ilişkilerinde bu sayede ortaya çıkartılması sağlanmıştır.

1.1 Tezin Amacı

Bu çalışmada Malatya'nın zirai meteoroloji haritası çıkartılarak üreticinin yapacağı yatırımı önceden planlayıp yanlış yatırımlardan doğacak maddi kayıpları en aza indirmesi ve yapılacak büyük projelerde ülke ekonomisine katkıda bulunulması amaçlanmıştır.

Uygulama olarak, milyonlarca bitlik verinin saklı olduğu meteoroloji ve tarımsal veri tabanları veri madenciliği modellerinden kümeleme ve sınıflandırma modelleriyle işlenip yorumlanması sağlanmış ve bunun sonucunda Malatya'nın zirai meteoroloji haritası çıkartılmıştır. Çıkartılan bu harita veri madenciliği modellerinden ilişkilendirme modeli ile işlenerek Doğu Anadolu Bölgesi tarım ürünlerinin desen keşfinde kullanılmıştır.

Kümeleme modeli ile ürünlerin yapısal özelliklerine göre hangi kümeye ait olduğu, sınıflandırma modeli ile ürünlerin verim analizlerine göre önceden belirlenmiş sınıflardan

hangisi ile temsil edileceđi, ilişkilendirme modeli ile de ürünler arasındaki gizli ilişkilerin (desenlerin) ortaya çıkarılması sağlanmıştır.

Sonuç olarak , tarım ürünün hangi kümeyi ait olduđu, bu küme içerisinde hangi sınıfla temsil edileceđi ve bu ürünün varlığıyla diđer yetiřecek ürünlerin varlıklarının tahmin edilmesi sağlanmıştır.

1.2. Tezin İçeriđi

Tez yukarda belirtilen amaçlara göre şöyle düzenlenmiştir; Veri madenciliđi süreci işlenerek Malatya'nın zîrai meteoroloji haritasının çıkartılması ve buna bađlı olarak tarımsal ürünlerin desen keřfinin yapılması sağlanmıştır.

Veri madenciliđi süreci problemin tanımlanması, verinin hazırlanması, modelin kurulması ve deđerlendirilmesi, modelin kullanılması ve modelin izlenmesi aşamalarını kapsamaktadır. Özellikle verinin hazırlanması aşaması, veri madenciliđinin en önemli aşamalarından biridir; çünkü sonuçların etkileri seçilen sayısal verinin dođruluđu ile orantılıdır. Veri hazırlama aşaması kendi içinde toplama, birleřtirme ve temizleme, dönüřtürme işlemlerini kapsamaktadır.

Toplama adımında, öncelikle uzun yıllar boyunca Malatya da yapılan meteoroloji uygulamalarında elde edilen verilerin saklandıđı 1. veri tabanı oluşturulmuştur. Bu veri tabanı meteoroloji uygulamaları sonucu elde edilen ortalama sıcaklık, ortalama minimum sıcaklık, ortalama maksimum sıcaklık, nisbi nem, güneşlenme süresi, yađış ortalaması ve rüzgar hızı verilerini içermektedir. Daha sonra tarımsal ürünlerin yetiřmesi için gerekli olan iklimsel özelliklerin saklandıđı 2. veri tabanı oluşturulmuştur. Veri tabanında tarım ürününün yetiřmesi için gerekli olan optimum (ortalama) sıcaklık, gerekli en düşük (min) sıcaklık, dayanabileceđi en yüksek (max) sıcaklık, gerekli nem deđer aralıđı, gerekli güneşlenme süresi, yıl içerisindeki toplam yađış miktarı, yıl içerisindeki toplam sıcaklık miktarını içermektedir.

Birleřtirme ve temizleme adımının ilk aşaması olan birleřtirme aşamasında farklı kaynaklardan alınarak toplanan verilerde bulunan farklılıklar giderilmiştir. Bu farklılıklar aynı deđer ifade eden başlıkların eşleşmesi anlamını taşımaktadır(Optimum sıcaklık=Ortalama sıcaklık, En düşük sıcaklık=Minimum sıcaklık gibi...). Temizleme aşamasında ürünlerin verimli olarak yetiřmesini sağlayan etkin ay aralıkları için 1 parametresi ürünün yetiřmesine etki etmeyen ay aralıkları için 0 parametresi tanımlanmıştır.

Dönüřtürme adımında verinin yapısıyla oynanmadan uygulanacak olan kümeleme ve sınıflandırma modelleri için gösterimi belirlenmiştir.

Modelin kurulması ve değerdendirilmesi aşamasında, veri madenciliğı modelleri kullanarak veri hazırlama aşamasında hazırlanan veriler üzerinde modellerin kurulmasını sağlar. Uygulamada, veri madenciliğı modellerinden kümeleme, sınıflandırma ve ilişkilendirme modelleri kullanılmıştır.

- Kümeleme: Veriler arasında benzer karakteristik değerdler taşıyan nesneleri bir araya getirerek gruplar oluşturmayı hedefler.
- Sınıflandırma: Verileri ait oldukları tanımlı sınıflara koymayı hedefler.
- İlişkilendirme: Veri tabanındaki veriler arasındaki ilişkileri bularak bir nesnenin(nesnelerin) varlığıyla diğerd bir nesnenin(nesnelerin) varlığını tahmin etmeyi hedefler.

Uygulamamızda her ürün içinde sık geğden terimler bulundu. Bulunan bu terimlerden ve ağırlıklarından yararlanılarak bir benzerlik ölçütü geliştirildi. Bu ölçüt ile ürünler yapısal özelliklerine göre 9 grupta kümelendirildi. Kümeleme modelinde, oluşan kümelerin kalitesi seçilen benzerlik ölçütüne ve bu ölçütün gerğekleşmesine bağıldır.

Uygulamada sınıflandırma yöntemlerinden karar ağağları (decision trees) tekniğı kullanıldı. Karar ağağları tekniğı ile istenilen örneğinin nitelikleri ağağ boyunca sınanır ve ulaşılan yaprağın etiketi sınıf bilgisini verir. Karar ağağları tekniğı ile ürünlerin verimsiz, vegatatif, az verimli ve verimli olarak önceden belirlenmiş sınıflardan hangisine dahil olduğı bulunmuştur. Bu teknik ile küçük ağağların yorumlanması kolaydır, ayrıca sürekli ve ayrık nitelik değerdleri için kullanılabilir.

Uygulamada ilişkilendirme yöntemlerinden apriori algoritması kullanarak bir nesnenin varlığı ile diğerd bir nesnenin varlığının tahmin edilebilmesi sağlanmıştır. Böylelikle Doğı Anadolu Bölgesi tarım ürünleri desen keşfinin yapılması sağlanmıştır.

Ancak dikkatten kaçmaması gereken bir diğerd ayrıntı veri madenciliğı kendi başına bir çözümd değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır.

Bu noktada karşımıza veri madenciliğı bir çözümd aracı olarak çıkmaktadır. Fakat madenciliğı yapılacak olan verinin de bazı vasıflara sahip olması gerekmektedir. Bu vasıflar veri ambarı ile sağlanmaktadır. Veri ambarları basit olarak veri madenciliğı işleminin yapılacağı verilerin oluşturulduğı özel veri tabanlarıdır.

2.VERİ MADENCİLİĞİ

Veri madenciliği, önceden bilinmeyen, geçerli ve uygulanabilir bilginin veri yığınlarından dinamik bir süreç ile elde edilmesi olarak tanımlanabilir.[1]

Bilgisayar sistemleri ile üretilen veriler tek başlarına değersizdirler, özellikle veri tabanlarının bilgiyi sadece saklamak için dizayn edildiği düşünüldüğünde çünkü çıplak gözle bakıldığında verilerin bir anlam ifade etmediğini söyleyebiliriz. Bu veriler belirli bir amaç doğrultusunda işlendiği zaman anlamlı hale gelmektedir. İşte ham veriyi bilgiye ve anlamlı hale dönüştürme işini veri madenciliği ile yapabiliriz.

Veri madenciliği kendi başına bir çözüm değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır. Veri madenciliği; analistciye, iş yapma aşamasında oluşan veriler arasındaki şablonları ve ilişkileri bulma sı konusunda yardım etmektedir.

Veri madenciliği, istatistik, matematik disiplinleri, modelleme teknikleri, veri tabanı teknolojisi ve çeşitli bilgisayar programları kullanılarak yapılmaktadır. Ayrıca bu süreçte kümeleme, veri özetleme sınıflama kurallarının öğrenilmesi, bağımlılık ağlarının bulunması, değişkenlik analizi ve anomali tespiti gibi farklı birçok teknik kullanılmaktadır.[3]

2.1. Veri Ambarının Oluşturulması

Veri ambarı veri madenciliğinin önemli bir aşamasıdır. Madenciliğini yapacağımız veri tek bir yapı içerisinde bulunmayabilir. Bu nedenle verinin tek bir çatı altında toplanması gerekir. Fakat veri ambarı oluşturma aşaması sadece verinin tek bir çatı altında toplanması değildir. Aynı zamanda toplanan veriler içerisinde var olan hataların ve belirsizliklerinde temizlenmesi aşamasıdır. Bu aşamada veri bazı alt işlemlere tabi tutulmaktadır. Bu işlemler Veri Toplama, Uyumlandırma, Birleştirme ve Temizlenme, Seçme ve Dönüştürmedir.

Toplama : Madenciliğini yapacağımız veri farklı kaynaklarda bulunabilir. Toplama işlemi, farklı kaynaklardan alınan bu verilerin tek bir kaynak altında birleştirilmesinin sağlandığı adımdır.

Uyumlandırma : Farklı kaynaklardan alınarak toplanan veri tipleri arasındaki uyumsuzlukların giderilmesinin sağlandığı adımdır. Bunun en yaygın örneği cinsiyette görülmektedir. Çok fazla tipte tutulabilen bir veri olup, bir veri tabanında 0/1 olarak tutulurken diğer veri tabanında E/K veya Erkek/Kadın şeklinde tutulabilir.

Birleştirme ve Temizleme : Uyumlandırma işlemi sırasında farklı veri kaynaklarından gelen verilerin birleştirilmesinin veya fazlalıkların temizlenmesinin sağlandığı adımdır.

Seme : İlerde kuracaėımız model iin uygun verinin seilmesinin saėlandığı adımdır. Veri tabanlarındaki ilem hızları artmasına raėmen büyük veritabanları üzerinde birden fazla modelin denenmesi olduka zaman ve maliyet gerektirmektedir. Bunun yerine verinin bütünü nü temsil edecek şekilde bir para üzerinde ilemler yapılabilir. Fakat seilecek paranın verinin tamamını temsil etmesi aısından önemi büyüktür.

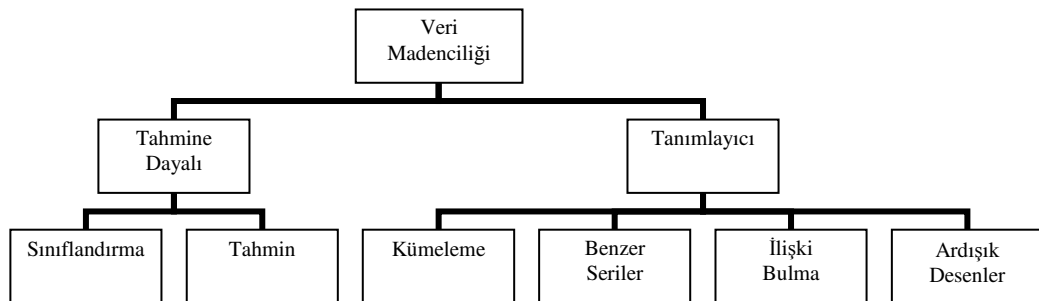
Dönüştürme : Verinin kullanılacak modele göre ieriğini koruyarak şeklini dönüştürülmesinin saėlandığı adımdır. Dönüştürme ilemi kullanılacak modele uygun biçimde yapılmalıdır. Çünkü verinin gösterilmesinde kullanılacak model ve algoritma önemli bir rol oynamaktadır.

2.2. Veri Madenciliėi Modelleri

Veri madenciliğinde kullanılan modeller, tahmin edici (Predictive) ve tanımlayıcı (Descriptive) olmak üzere iki ana başlık altında toplanabilir.[4]

Tahmin edici model, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç deėerlerin tahmin edilmesi amaçlanmaktadır. Örneėin bir banka önceki dönemlerde vermiş olduėu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu verilerde baėımsız deėişkenler kredi alan müşterinin özellikleri, baėımlı deėişken deėeri ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır.

Tanımlayıcı model, karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması saėlanmaktadır. 25 yaşı altı bekar kişiler ile, 25 yaşı üstü evli kişiler üzerinde yapılan ve ödeme performanslarını gösteren bir analiz tanımlayıcı modellere örnek olarak verilebilir.



Şekil 2. 1. Veri Madenciliėi Modelleri

2.3. Veri Madenciliğinde Kullanılan Algoritmalar

Veri madenciliği; veri kaynağının seçimi, analiz edilecek kayıtların bazı dönüşümlerle hazırlanması, uygun bir veri madenciliği algoritmasının çalıştırılması, sonuçların değerlendirilmesi süreçlerinden oluşan ve geçerli sonuca ulaşılan kadar kendini tekrarlayan dinamik bir süreçtir. Veri madenciliği algoritmaları kayıtların içindeki gizli ilişkileri ortaya çıkaran matematiksel tekniklerdir.

2.3.1. İlişki Bulma Algoritması

İlişki bulma algoritmasının amacı bir işlem içindeki bazı kalemlerin bir arada olmasının rastlantısal olmadığı durumları bulabilmektir. Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri bulurlar.[8] Aynı küme içinde bulunan elemanlar arasındaki ilişkileri ortaya çıkarır. Örnek olarak satış kayıtlarının olduğu bir veritabanını ele alalım. İlişki Bulma Algoritması ile birlikte satın alınan ürünleri bulabiliriz. Kartpostal alan müşterilerin %60'ının kozmetik ürünlerinden de aldığı şeklinde bir sonuç elde edebilirsiniz.

İlişkilendirme kurallarını bulma:

- Bütün öğelerden oluşan küme $I = \{i_1, i_2, \dots, i_d\}$
 - $I = \{\text{ekmek, süt, bira, kola, yumurta, bez}\}$
- Hareket $T_j \subseteq I$, $T_j = \{i_{j1}, i_{j2}, \dots, i_{jk}\}$
 - $T1 = \{\text{ekmek, süt}\}$
- Hareketlerden oluşan veri kümesi $D = \{T_1, T_2, \dots, T_N\}$ (Tablo2.1)

Tablo 2.1. Market sepet verisi

TID	Öğeler
T1	Ekmek, Süt
T2	Ekmek, Bez, Bira, Yumurta
T3	Süt, Bez, Bira, Kola
T4	Ekmek, Süt, Bez, Bira
T5	Ekmek, Süt, Bez, Kola

Yaygın Öğeler:

Bez, bira

Süt, ekmek, yumurta, kola

Bira, ekmek, süt

Bulunan İlişkilendirme Kuralları

$\{\text{Bez}\} \rightarrow \{\text{Bira}\}$

$\{\text{Süt, Ekmek}\} \rightarrow \{\text{Yumurta, Kola}\}$

$\{\text{Bira, Ekmek}\} \rightarrow \{\text{Süt}\}$

Yaygın Öğeler :

- Öğeler kümesi (Itemset)

- Bir veya daha çok öğeden oluşan küme
- k-öğeler kümesi (k-itemset): k öğeden oluşan küme

3-öğeler kümesi: {Bez, Bira, Ekmek} (Tablo2.1)

- Destek sayısı σ (Support count)

- Bir öğeler kümesinin veri kümesinde görülme sıklığı

$\sigma(\{\text{Süt, Ekmek, Bez}\}) = 2$ (Tablo2.1)

- Destek s (Support)

- Bir öğeler kümesinin içinde bulunduğu hareketlerin toplam hareketlere oranı

$s(\{\text{Süt, Ekmek, Bez}\}) = 2/5$ (Tablo2.1)

- Yaygın öğeler (Frequent itemset)

- Destek değeri minsup eşik değerinden daha büyük ya da eşit olan öğeler kümesi

İlişkilendirme Kuralları :

- Veri kümesi $D=\{T_1, T_2, \dots, T_N\}$

- En az, en küçük destek ve güven değerine sahip $X \rightarrow Y$ şeklinde kuralların bulunması denklem (2.1)'de verilmiştir.

- $X \subset I, Y \subset I, X \cap Y = \emptyset$ (2.1)

- Kuralları değerlendirme ölçütleri

- destek (support) s: $X \cup Y$ öğeler kümesinin bulunduğu hareket sayısının toplam hareket sayısına oranı denklem (2.2)'de verilmiştir.

$$\text{sup port}(X \rightarrow Y) = \frac{\# X \cup Y}{N} \quad (2.2)$$

- güven (confidence) c: $X \cup Y$ öğeler kümesinin bulunduğu hareket sayısının X öğeler kümesi bulunan hareket sayısına oranı denklem (2.3)'de verilmiştir.

$$\text{confidence}(X \rightarrow Y) = \frac{\# X \cup Y}{\# X} \quad (2.3)$$

Örnek :

$\{\text{Süt, Bez}\} \rightarrow \{\text{Bira}\}$

$$\text{sup port} = \frac{\sigma(\text{süt, bira, bez})}{|T|} = \frac{2}{5} = 0,4 \quad (\text{Tablo2.1})$$

$$\text{confidence} = \frac{\sigma(\text{süt, bira, bez})}{\sigma(\text{süt, bez})} = \frac{2}{3} = 0,67 \quad (\text{Tablo2.1})$$

- İlişkilendirme kuralları madenciliğinde temel amaç D hareket kümesinden kurallar oluşturmak

- kuralların destek değeri, belirlenen en küçük destek (minsup) değerinden büyük yada eşit olmalı
- kuralların güven değeri, belirlenen en küçük güven (minconf) değerinden büyük yada eşit olmalı

- Brute-force yaklaşım

- Olası bütün kuralları listele
- Her kural için destek ve güven değeri hesapla
- minsup ve minconf eşik değerlerinden küçük destek ve güven değerlerine sahip kuralları sil
- hesaplama maliyeti yüksek

İlişkilendirme Kuralları Oluşturma:

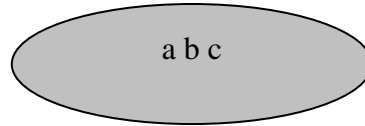
İki adımda gerçekleşir.

1. Şekil 2.2’de yaygın öğeleri belirleme :

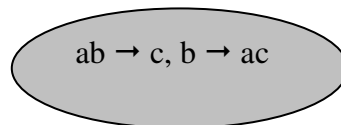
- destek değeri minsup değerinden büyük ya da eşit olan öğeler kümelerini belirlemek

2. Şekil 2.3’de kural Oluşturma :

- Destek değeri minconf değerinden büyük ya da eşit olan ve yaygın öğelerin ikili bölünmeleri olan kurallar oluşturma
- Güçlü kurallar



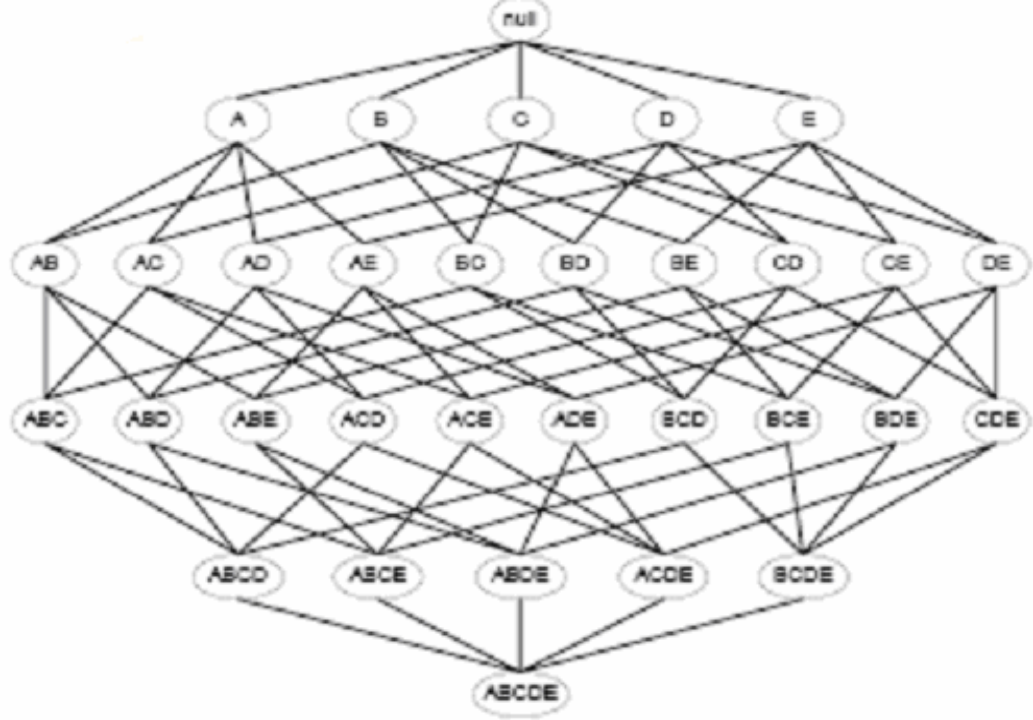
Şekil 2.2. Yaygın öğeleri belirleme



Şekil 2.3. İlişkilendirme kuralları oluşturma

Yaygın Öge Adayları Oluşturma:

Yaygın öğeleri bulma maliyeti; d öge için $2^d - 1$ yaygın öge oluşabilir



Şekil 2. 4. Yaygın öge adaylarının bulunması

Apriori Algoritması:

- Bir yaygın ögenin herhangi bir altkümesi de yaygın ögedir.

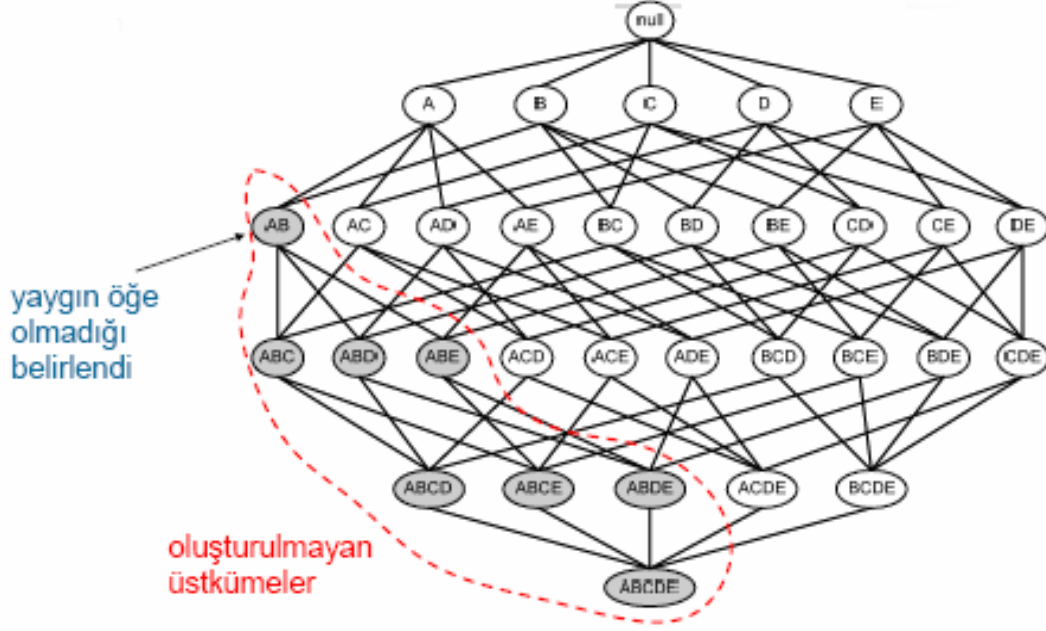
- {Süt, Bez, Bira} kümesi yaygın öge ise {Süt, Bez} kümesi de yaygın ögedir
- {Süt, Bez, Bira} öğeler kümesi olan her harekette {Süt, Bez} kümesi de vardır

- Yaygın öge aday sayısını azaltma yöntemi: Yaygın öge olmayan bir kümenin üst kümeleri yaygın öge adayı olarak oluşturulmaz (destek değeri hesaplanmaz)

- Yöntem:

- k-yaygın öğeler kümesinden (k+1) yaygın öge adayları oluştur
- yaygın öge adayları için destek değeri hesapla

Apriori Yaygın Öğe Azaltma Yöntemi :



Şekil 2. 5.Yaygın öğe azaltma yöntemi

2.3.2. Sınıflandırma Algoritması

Sınıflandırma algoritması, daha önce ayırmış veriler üzerinde bir model kurar. Bu modeli daha önce yapılan ayrıştırmanın temellerini açıklamak veya yeni verileri sınıflandırmak amacı ile kullanabiliriz. Bu yolla geçmişe ait verilerden yola çıkarak ileriye dönük tahminler yapılabilir.

Örneğin bir sigorta şirketi geçmişte sigortalarını devam ettirmeyen müşterilerinin özelliklerini analiz ederek bir risk grubu profili oluşturabilir. Bu profil model olarak kullanılarak yeni müşterilerin risk grubuna dahil olup olmadığı test edilebilir.

- Öğrenme modunda, geçmişe ait veriler üstünde çalışarak önceden tanımlanmış risk gruplarının özellikleri öğrenilir.
- Test modunda , yaratılan modelin doğruluğu geçmişe ait bir başka veri setinde test edilir.
- Uygulama modunda, doğruluğu ispatlanan model yeni müşterilerin risk gruplarını tahmin etmekte kullanılır.

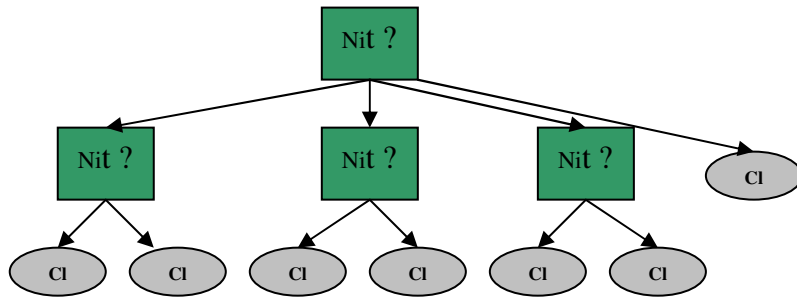
Aynı teknik sigortacılıkta hasar karşılama taleplerinin cevaplanmasında, kredi kartı yolsuzluklarının tespitinde, imalat sanayi kalite kontrol sistemlerinde ve buna benzer hata durumlarının kontrolünde kullanılabilir. Diğer uygulama alanları arasında pazarlama kampanyaları, tıpta teşhis ve tedaviye yanıt ölçümü, stok yönetimi ve raf yerleşimi planlaması sayılabilir. Temel sınıflandırma yöntemleri[9]:

- Karar ağaçları
- Yapay sinir ağları
- Bayes sınıflandırıcı
- Bayes ağları

2.3.2.1. Karar Ağaçları ile Sınıflandırma

Akış diyagramı şeklinde ağaç yapısı şekil 2.6’da gösterilmiştir.

- Ara düğümler, bir nitelik sınavasını **Nit ?**
- Dallar, sınama sonucunu →
- Yapraklar, Sınıfları temsil eder. **Cl**



Şekil 2.6. Karar ağaç yapısı

Karar Ağacı Yöntemleri :

Karar ağaçları algoritmaları verinin belirli sınıflara atanması için kullanılır.[7] Karar ağacı oluşturma yöntemleri genel olarak iki aşamadan oluşur:

- Ağaç Oluşturma

- En başta bütün öğrenme kümesi örnekleri kökte
- Seçilen niteliklere bağlı olarak örnek yinelenmeli olarak bölünüyor

- Ağaç Budama

- Öğrenme kümesindeki gürültülü verilerden oluşan ve sına kümesinde hataya neden olan dalları silme (sınıflandırma başarımını artırır).

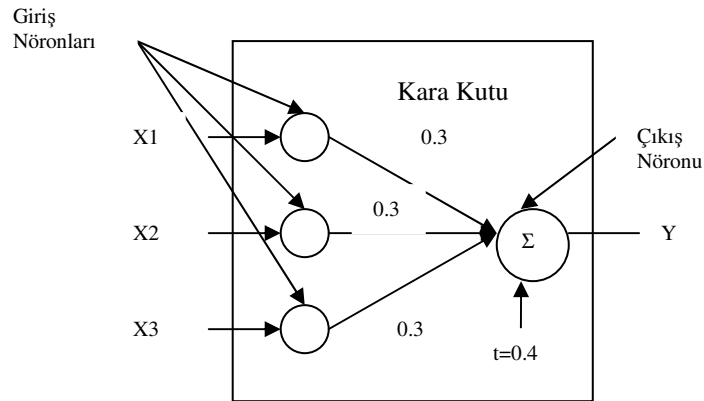
2.3.2.2. Yapay Sinir Ağları ile Sınıflandırma

- İnsan beynindeki sinir hücrelerinin işlevini modelleyen bir yapı
- Birbiri ile bağlantılı katmanlardan oluşur.
 - katmanlar hücrelerden oluşur
- Katmanlar arasında iletim
- İlet katmanlar arasındaki bağın ağırlığına ve her hücrenin değerine bağlı olarak değişebilir.

Örnek:

Tablo 2.2. Yapay sinir ağı uygulaması

X1	X2	X3	Y
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	1
0	0	0	0



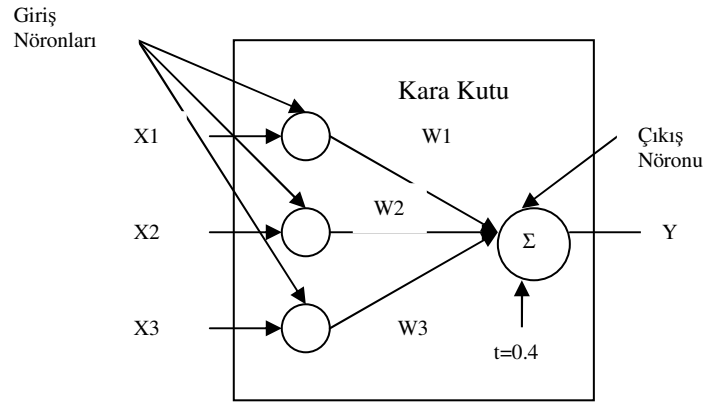
Şekil 2.7. Yapay sinir ağı uygulaması

$$Y=I(0.3X_1 + 0.3X_2+ 0.3X_3 - 0.4 >0)$$

$$I(Z)= \begin{cases} 1 & \text{eğer } z>0 \\ 0 & \text{diğer} \end{cases} \quad (2.4)$$

Yapay sinir ağıları :

- Birbiri ile bağlantılı nöronlar ve ağırlıklar
- Çıkış nöronu kendisine gelen girişleri ağırlıklı olarak denklem (2.5)'de topluyor.
- Çıkış nöronu bir eşik değeri ile karşılaştırılıyor.



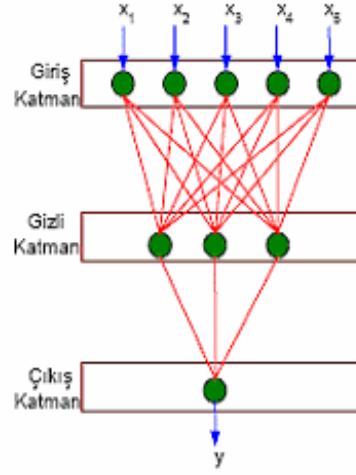
Şekil 2.8. Yapay sinir ağı

$$Y = I\left(\sum_i w_i X_i - t\right) \quad (2.5)$$

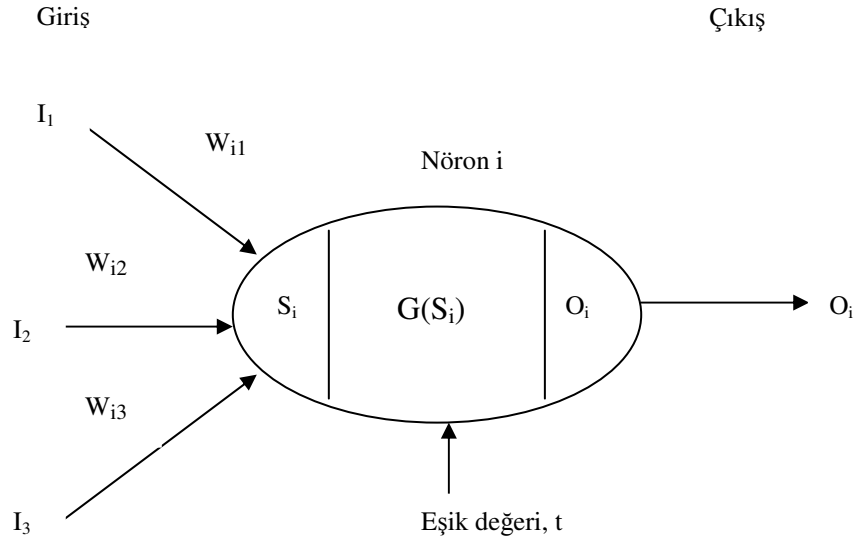
$$Y = \text{sign} \left(\sum_i w_i X_i - t \right)$$

Yapay Sinir Ağını Oluşturma:

- Giriş nöron sayısı şekil 2.9'da verilmiştir.
 - Öğrenme kümesindeki verilerin nitelik sayısı
- Gizli nöron sayısı şekil 2.9'da verilmiştir.
 - Öğrenme sırasında ayarlanır
- Çıkış nöron sayısı şekil 2.9'da verilmiştir.
 - Sınıf sayısı



Şekil 2.9. YSA oluşturma



Şekil 2.10. YSA ağırlıklarını öğrenme

2.3.2.3. Bayes Teoremi ile Sınıflandırma

- X sınıflandırılacak örnek. Hipotez h, X örneğinin C sınıfına ait olduğu
- h hipotezinin sonrasal olasılığı denklem (2.6)'da verilmiştir.

$$P(h|X) = \frac{P(X|h)P(h)}{P(X)} \quad (2.6)$$

- MAP (maximum posteriori) hipotez denklem (2.7)'de verilmiştir.

$$h_{\text{MAP}} \equiv \arg \max_{h \in H} P(hD) = \arg \max_{h \in H} P(Dh)P(h) \quad (2.7)$$

- Çok sayıda olasılığı önceden kestirmek gerekiyor

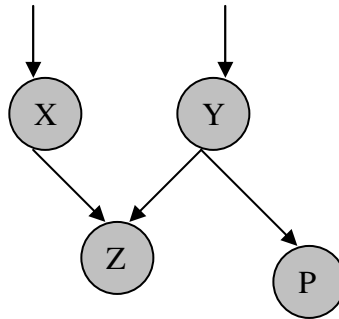
2.3.2.4. Bayes Ağları ile Sınıflandırma

- Niteliklerin altkümesinin birbiri ile bağımsız olduğunu varsayıyor

- Yönlü çevrimsiz çizge (directed acyclic graph) ve koşullu olasılık tablolarından oluşur

- Her değişken A için bir tablo var

- niteliğin ebeveynlerine olan koşullu olasılıkları
- Düğümler: rasgele değişkenler
- Ayrıklar: olasılıklı bağılılık
- X ve Y, Z değişkeninin ebeveyni
- Y, P değişkeninin ebeveyni
- Z ve P arasında bağ yok



Şekil 2.11. Bayes ağları

2.3.2.5. Modellerin Değerlendirmesi

- Modelin başarımı (doğruluğu) sınama kümesi örnekleri kullanılarak belirlenir.

- Sınıf etiketi bilinen bir sınama kümesi örneği model kullanılarak belirlenen sınıf etiketiyle karşılaştırılır.

- Modelin doğruluğu, doğru sınıflandırılmış sınama kümesi örneklerinin toplam sınama kümesi örneklerine oranı olarak belirlenir.

- Sınama kümesi model öğrenirken kullanılmaz.

2.3.3. Kümeleştirme Algoritması

Kümeleştirmenin amacı kayıtları ortak özellikleri olanlar bir arada olacak şekilde gruplandırmaktır. Kayıtların en sık görülen ortak özelliklerinin belirlenmesi ve bu ortak özelliklere göre kayıt gruplarının oluşturulması. Sonuç olarak bulunan gruplar, her grubun belirleyici özellikleri ve bu özelliklerin grup içinde dağılımı elde edilir. Örnek olarak bir süpermarketin müşteri bilgileri ve satış kayıtları kümeleştirme algoritması ile incelenecek olursa, müşterilerin büyük bir kısmının düzenli olarak Cuma akşamları kredi kartıyla alışveriş yaptıkları şeklinde bir sonuca ulaşılabilir.

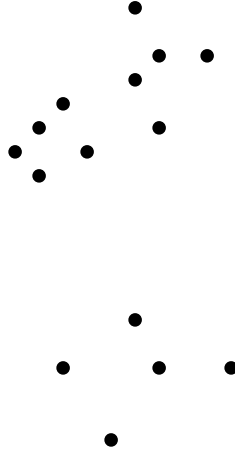
Kümeleme analizinde veri tabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklere göre yapılacağı konunun uzmanı olan bir kişi tarafından yapılabileceği gibi veri tabanındaki kayıtların hangi kümelere ayrılacağını geliştirilen bilgisayar programları da yapabilmektedir. Temel Kümeleme Yaklaşımları:[5]

- Bölünmeli yöntemler: Veriyi bölerek, her grubu belirlenmiş bir kritere göre değerlendirir.
- Hiyerarşik yöntemler: Veri kümelerini (yada nesneleri) önceden belirlenmiş bir kritere göre hiyerarşik olarak ayırır.
- Yoğunluk tabanlı yöntemler: Nesnelerin yoğunluğuna göre demetleri oluşturur.
- Model tabanlı yöntemler: Her demetin bir modele uyduğu varsayılır. Amaç bu modellere uyan verileri gruplamak.

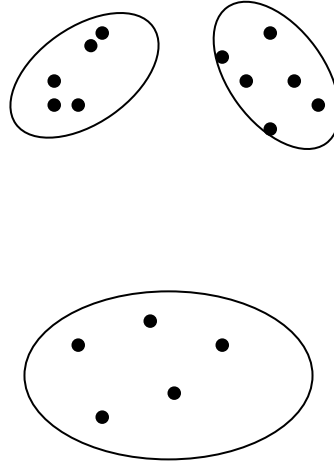
2.3.3.1. Bölünmeli Yöntemler ile Kümeleme

- Amaç: Şekil 2.12'deki n nesneden oluşan bir veri kümesini (D) k ($k \leq n$) demete ayırmak.
- Şekil 2.13'de her demette en az bir nesne bulunmalı.
- Şekil 2.13'de her nesne sadece bir demette bulunmalı.
- Yöntem: Kümeleme kriterini en büyütecek şekilde D veri kümesi k gruba ayırma.
- Global çözüm: Mümkün olan tüm gruplamaları yaparak en iyisini seçme (NP karmaşık).
- Sezgisel çözüm: k -means ve k -medoids.

Bölünmeli Denetleme:



Şekil 2.12. Veri kümesi



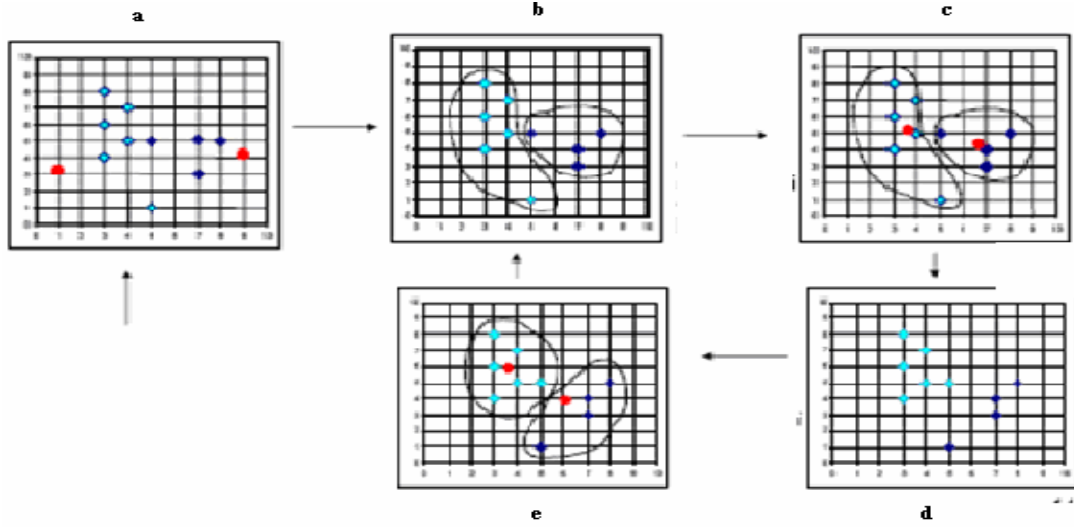
Şekil 2.13. Bölünmeli demetleme

K-Metod ile Demetleme:

- Bilinen bir k değeri için k-means demetleme algoritmasının 4 aşaması vardır:

1. Veri kümesi k altkümeye ayrılır (her demet bir altküme).
2. Her demetin ortalaması hesaplanır: merkez nokta
3. Her nesne en yakın merkez noktanın olduğu demete dahil edilir.
4. Nesnelerin demetlenmesinde değişiklik olmayana kadar adım 2'ye geri dönülür.

Örnek:



Şekil 2.14. K Metod uygulaması

K-Means Demetleme Yöntemini Değerlendirme:

- Yaygın olarak kullanılan yöntem hataların karelerinin toplamı (Sum of Squared Error SSE).

- Nesnelerin bulundukları demetin merkez noktalarına olan uzaklıklarının karelerinin toplamı denklem (2.8)'de verilmiştir.

$$SSE = \sum_{l=1}^K \sum_{X \in C_l} dist^2(m_l, X) \quad (2.8)$$

x: Ci demetinde bulunan bir nesne, m_i : Ci demetinin merkez noktası

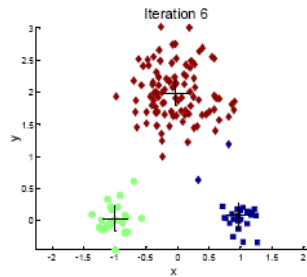
- Hataların karelerinin toplamını azaltmak için k demet sayısı artırılabilir.

- Küçük k ile iyi bir demetleme, büyük k ile kötü bir demetlemeden daha az SSE değerine sahip olabilir.

- Başlangıç için farklı merkez noktaları seçerek farklı demetlemeler oluşturulur.

- En az SSE değerini sahip olan demetleme seçilir.

Merkez Noktaların Seçimi:



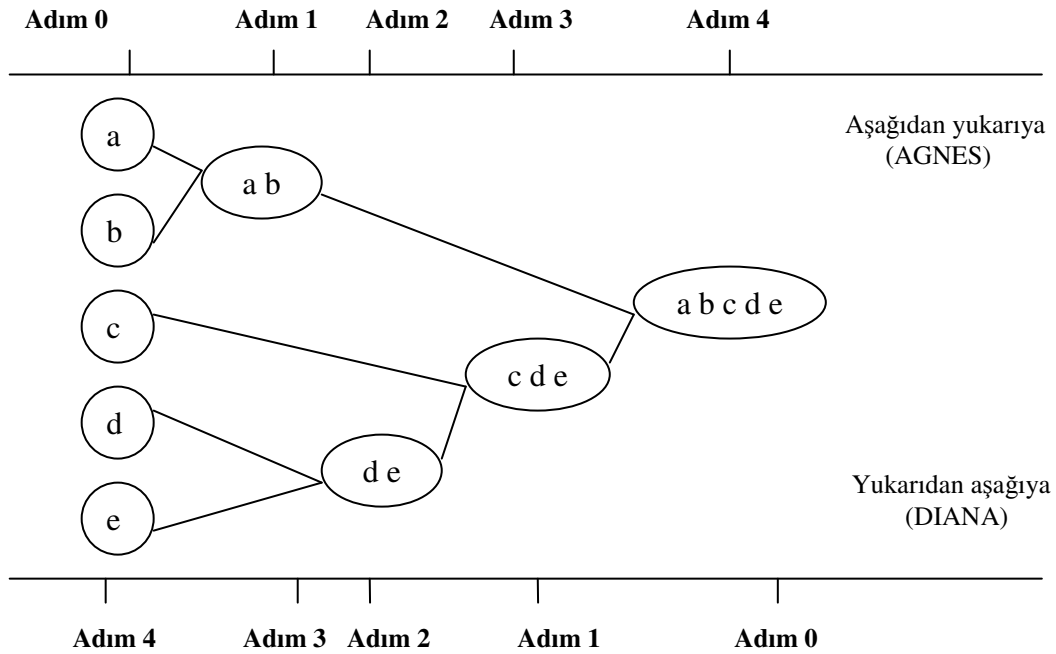
Şekil 2.15. Merkez nokta seçimi

K-Means Demetleme Algoritmasının Özellikleri :

- Gerçeklemesi kolay
- Karmaşıklığı diğer demetleme yöntemlerine göre az
- K-Means algoritması bazı durumlarda iyi sonuç vermeyebilir
 - Veri grupları farklı boyutlarda ise
 - Veri gruplarının yoğunlukları farklı ise
 - Veri gruplarının şekli küresel değilse
 - Veri içinde aykırılıklar varsa

2.3.3.2. Hiyerarşik Yöntemler ile Kümeleme

- Demet sayısının belirlenmesine gerek yok
- Sonlanma kriteri belirlenmesi gerekiyor



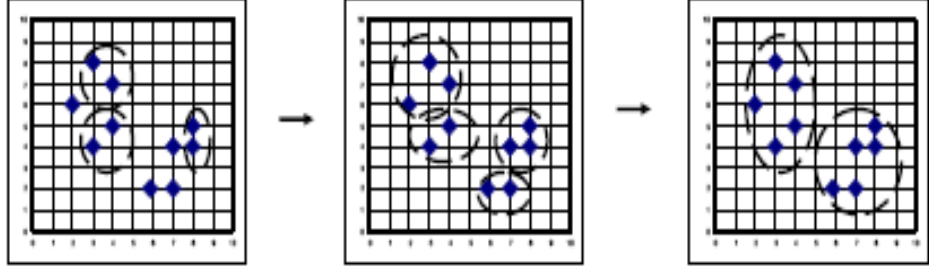
Şekil 2.16. Hiyerarşik yöntem

Hiyerarşik Yöntemler :

- AGNES (AGglomerative NESTing):

- Kaufmann ve Rousseeuw tarafından 1990 yılında önerilmiştir.
- Birinci adımda her nesne bir demet oluşturur.
- Aralarında en az uzaklık bulunan demetler her adımda birleştirilir.

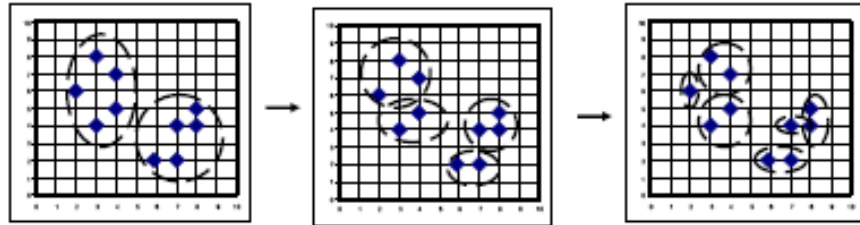
- Bütün nesneler tek bir demet içinde kalana kadar ya da istenen sayıda demet elde edene kadar birleştirme işlemi devam eder.



Şekil 2.17. Aşağıdan yukarıya doğru kümeleme (AGNES)

DIANA (DIvisive ANALysis):

- Kaufmann ve Rousseeuw tarafından 1990 yılında önerilmiştir.
- AGNES'in yaptığı işlemlerin tersini yapar.
- En sonunda her nesne bir demet oluşturur.
- Her nesne ayrı bir demet oluşturan ya da istenilen demet sayısı elde edene kadar ayrılma işlemi devam eder.[6]



Şekil 2.18. Yukarıdan aşağıya doğru kümeleme (DIANA)

Hiyerarşik Demetleme Yöntemlerinin Özellikleri :

- Demetleme kriteri yok.
- Demet sayılarının belirlenmesine gerek yok.
- Aykırılıklardan ve hatalı verilerden etkilenir.
- Farklı boyuttaki demetleri oluşturmak problemli olabilir.
- Yer karmaşıklığı – $O(n^2)$.
- Zaman karmaşıklığı – $O(n^2 \log n)$. n : nesne sayısı.

2.3.3.3. Yoğunluk Tabanlı Yöntemler ile Kümeleme

- Demetleme nesnelerin yoğunluğuna göre yapılır[10].

- Başlıca özellikleri:

- Rasgele şekillerde demetler üretilebilir.
- Aykırı nesnelerden etkilenmez.
- Algoritmanın son bulması için yoğunluk parametresinin verilmesi gerekir.

- Başlıca yoğunluk tabanlı yöntemler:

- DBSCAN
- OPTICS
- DENCLUE
- CLIQUE

DBSCAN

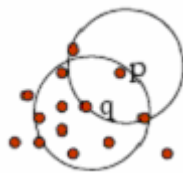
- İki parametre:

- Eps: En büyük komşuluk yarıçapı
- MinPts: Eps yarıçaplı komşuluk bölgesinde bulunan en az nesne sayısı

- $Neps(p)$: $\{q \in D \mid d(p,q) \leq Eps\}$

- Doğrudan erişilebilir nesne: Eps ve MinPts koşulları altında bir q nesnesinin doğrudan erişilebilir bir p nesnesi şu şartları sağlar:

- $p \in Neps(q)$
- q nesnesinin çekirdek nesne koşulunu sağlaması şekil 2.19’da gösterilmiştir.



Şekil 2.19. DBSCAN

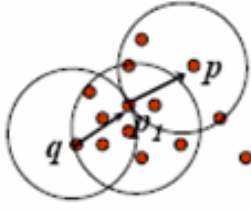
$$Neps(q) \geq MinPts$$

$$MinPts = 5$$

$$Eps = 1 \text{ cm}$$

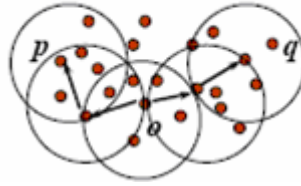
- Şekil 2.20’de erişilebilir nesne :

- Eps ve MinPts koşulları altında q nesnesinin erişilebilir bir p nesnesi olması için:
 - $p_1, p_2, \dots, p_n$ nesne zinciri olması,
 - $p_1 = q, p_n = p$,
 - p_i nesnesinin doğrudan erişilebilir nesnesi: p_{i+1}



Şekil 2.20. Erişebilir nesne

- Şekil 2.21’de yoğunluk bağlantılı Nesne :
- Eps ve MinPts koşulları altında q nesnesinin yoğunluk bağlantılı nesnesi p şu koşulları sağlar:
 - p ve q nesneleri Eps ve MinPts koşulları altında bir o nesnesinin erişilebilir nesnesidir.



Şekil 2.21. Yoğunluk bağlantılı nesne

2.3.3.4. Model Tabanlı Yöntemler ile Kümeleme

- İstatistiksel yaklaşım:

- K nesneden oluşan bir veri kümesi $D=\{x_1, x_2, \dots, x_K\}$
- Her x_i ($i \in [1, \dots, K]$) nesnesi Θ parametre kümesiyle tanımlanan bir olasılık dağılımından oluşturulur.
- Olasılık dağılımının, $c_j \in C=\{c_1, c_2, \dots, c_G\}$ şeklinde G adet bileşeni vardır.
- Her Θ_g , $g \in [1, \dots, G]$ parametre kümesi g bileşeninin olasılık dağılımını belirleyen, Θ kümesinin ayrışık bir alt kümesidir.
- Herhangi bir x_i nesnesi öncelikle, $p(c_g|\Theta) = \tau_g$, ($\sum_G \tau_g=1$ olacak şekilde) bileşen katsayısına (ya da bileşenin seçilme olasılığına) göre bir bileşene atanır.
- Bu bileşen $p(x_{ilc_g}|\Theta_g)$ olasılık dağılımına göre x_i değişkenini oluşturur.
- Böylece bir x_i nesnesinin bu model için olasılığı denklem (2.9)’da bütün bileşenlerin olasılıklarının toplamıyla ifade edilebilir.

$$p(x_i | \Theta) = \sum_{g=1}^G p(c_g | \Theta) p(x_i | c_g; \Theta_g) \quad (2.9)$$

$$p(x_i | \Theta) = \sum_{g=1}^G \tau_g p(x_i | c_g; \Theta_g)$$

Model Tabanlı Demetleme Problemi :

- Model parametrelerinin belirlenmesi

- Maximum Likelihood (ML) yaklaşımı

$$l_{ML}(\Theta_1, \dots, \Theta_G; \tau_1, \dots, \tau_G | D) = \prod_{i=1}^K \sum_{g=1}^G \tau_g p(x_i | c_g, \Theta_g) \quad (2.10)$$

- Maximum Aposteriori (MAP) yaklaşımı

$$l_{MAP}(\Theta_1, \dots, \Theta_G; \tau_1, \dots, \tau_G | D) = \prod_{i=1}^K \sum_{g=1}^G \frac{\tau_g p(x_i | c_g, \Theta_g) p(\Theta)}{p(D)} \quad (2.11)$$

- Uygulamada her ikisinin logaritması

$$L(\Theta_1, \dots, \Theta_G; \tau_1, \dots, \tau_G | D) = \prod_{i=1}^K \ln \sum_{g=1}^G (\tau_g p(x_i | c_g, \Theta_g)) \quad (2.12)$$

$$L(\Theta_1, \dots, \Theta_G; \tau_1, \dots, \tau_G | D) = \prod_{i=1}^K \ln \sum_{g=1}^G (\tau_g p(x_i | c_g, \Theta_g)) + \ln p(\Theta) \quad (2.13)$$

2.3.4. Ardışık Desenler Algoritması

Ardışık desenler algoritmasının amacı zaman içinde önceden tahmin edilebilir olay serileri bulmaktır. Başka bir deyişle algoritma belirli bir zaman aralığında gerçekleşen olaylar arasında sebep sonuç ilişkileri çıkarır. Mesela bankacılık verilerinde çalıştırılan bir ardışık desen algoritması yeni hesap açtıran müşterilerin %42'sinin 90 gün içinde ATM kartı için başvuruda bulunduğu sonucunu keşfedebilir. Bu algoritmayla sigortacılıkta hileli hasar ödeme talepleri tespit edilebilir; teçhizat yenilemesi, satış kampanyaları gibi aktiviteler planlanabilir.

2.3.5. Benzer Seriler Algoritması

Benzer seriler algoritması, serilerin (sayısal değer dizileri) tutulduğu bir veritabanında yaptığı analizle seriler içinde benzeşen kısımları (alt serileri) tespit eder. Bu algoritma tedarik zinciri yönetiminde (supply chain management) planlama amacıyla kullanılabilir.

Örneğin bir süpermarket satış verilerinde yaptığı bir çalışmayla mevsimsel talep eğrileri benzeşen ürünleri tespit edip, bu ürünler için ortak bir satın alım ve depolama planı

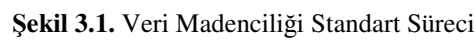
yapabilir. Aynı teknikle büyüme tipleri benzeşen şirketleri ve fiyat hareketleri benzeşen hisse senetlerini bulmak mümkündür. Diğer kullanım alanları arasında sismik dalgaların incelenmesi yoluyla yapılan deprem araştırmaları ve jeolojik farklılıkların tespiti sayılabilir.

2.3.6.Tahmin Algoritması

Tahmin algoritmasının amacı bir kayda bağlı özelliklerden birinin değeriyle diğer özelliklerin değerleri arasında bağımlılık olup olmadığını belirlemektir. Böyle bir bağımlılık görüldüğünde yeni kayıtlar için bağımlı olan özelliğin değerini tahmin edebilecek bir model oluşturulur.

Örneğin bir sigorta şirketi yeni müşterilerinden elde edeceği satış gelirini tahmin etmek istediğinde geçmişteki satış kayıtlarını müşteri özellikleriyle birlikte analiz ederek bir model oluşturabilir. Bu model kullanılarak yeni müşteriler için beklenen satış miktarları hesaplanabilir.

Ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda fayda sağlaması mümkün değildir. Bu nedenle yukarıda tanımlanan tüm aşamalardan önce, iş ve veri özelliklerinin öğrenilmesi / anlaşılması başarının ilk şartı olacaktır.



1. Problemin Tanımlanması,
2. Verilerin Hazırlanması,
3. Modelin Kurulması ve Değerlendirilmesi,
4. Modelin Kullanılması,
5. Modelin İzlenmesi

3.1. Problemin Tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağına açık bir şekilde tanımlanmasıdır. İlgili işletme amacı işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir.

Bu aşamada mevcut iş probleminin nasıl bir sonuç üretilmesi durumunda çözüleceğinin, üretilen sonucun fayda - maliyet analizinin başka bir değişle üretilen bilginin işletme için değerinin doğru analiz edilmesi gerekmektedir.

3.2. Verilerin Hazırlanması

Veri madenciliğinin en önemli aşamalarından bir tanesi olan verinin hazırlanması aşaması analistin toplam zaman ve enerjisinin %50 - %75 ini harcamasına neden olmaktadır. Bu aşamada firmanın mevcut bilgi sistemleri üzerinde ürettiği sayısal bilginin iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği unutulmamalıdır. Proje kapsamında kullanılacak sayısal verilerin, hangi iş süreçleri ile yaratıldığı da bu veriler kullanılmadan analiz edilmelidir, bu sayede analist veri kalitesi hakkında fikir sahibi olabilir. Verilerin hazırlanması aşaması kendi içerisinde toplama, birleştirme ve temizleme, dönüştürme adımlarından meydana gelmektedir.

3.2.1. Toplama

Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından faydalanılabilir.

3.2.2. Birleştirme ve Temizleme

Bu adımda toplanan verilerde bulunan farklılıklar giderilmeye çalışılır. Hatalı veya analizin yanlış yönlendirilmesine sebep olabilecek verilerin temizlenmesine çalışılır. Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

3.2.3. Dönüştürme

Kullanılacak model ve algoritma çerçevesinde verilerin tanımlama veya gösterim şeklinin de değiştirilmesi gerekebilir. Örneğin kredi riski uygulamasında iş tiplerinin, gelir seviyesi ve yaş gibi değişkenlerin kodlanarak gruplanmasının faydalı olacağı bilinmelidir.

3.3. Modelin Kurulması ve Değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (Simple Validation) testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır. (Doğruluk Oranı = 1 - Hata Oranı)

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, bir çok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir. Çok ender olarak yorumlanamayacak kadar karmaşıklaşsalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir.

3.4. Modelin Kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilmesi gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir.

3.5. Modelin İzlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir.

4. MALATYANIN ZİRAİ METEOROLOJİ HARİTASININ ÇIKARILMASI

4.1. Problemin Tanımlanması

Günümüzde tüm meteorolojik cihazların dijital hale gelmesi bu alanda aşırı veri birikmesine neden olmuştur. Amacımız kağıt üzerinde veri toplanan klasik meteoroloji bilgi sisteminden farklı olarak veri tabanında kayıtlı olan büyük veri yığınlarında gizli, önemli, önceden bilinmeyen ve yararlı bilgileri bulmaktır.

Bulunan bu bilgiler ile Malatya'nın zirai meteoroloji haritasının oluşturulması, oluşturulan bu harita ile üreticinin yapacağı yatırımı önceden planlaması böylelikle yatırımdan maksimum faydanın sağlanması yanlış planlamadan dolayı oluşacak zararın minimuma indirilmesi bu bağlamda yapılacak büyük projelerde ülke ekonomisine katkıda bulunulması sağlanacaktır.

Milyonlarca bitlik verinin saklı olduğu meteoroloji ve tarımsal veri tabanlarının bireysel çalışmalar ile işlenmesi ve yorumlanması neredeyse imkansızdır. Bu aşamada yardımımıza veri madenciliği teknikleri yetişmektedir. Böylelikle büyük veri yığınlarının işlenip yorumlanması mümkün hale gelmiş klasik yöntemlerle bulunması çok zor veya imkansız olan ilişkilerde ortaya çıkarılmıştır.

Bu çalışmada veri madenciliği tekniklerinde kümeleme, sınıflandırma ve ilişki bulma modelleri kullanılmıştır. Meteorolojik parametreler yardımıyla kümeleme tekniğinde tarımsal ürünler yapısal özelliklerine göre benzer familyalar bir arada olacak şekilde kümelenmiştir. Sınıflandırma tekniği ile kümelenen ürünlerin verimli, az verimli, vegetatif ve verimsiz olarak önceden belirlenmiş sınıflara dahil edilmesi sağlanmıştır. İlişki bulma tekniği ile bir nesnenin varlığı ile diğer bir nesnenin varlığının tahmin edilebilmesi sağlanmıştır.

Çalışmanın başarısı insanlar tarafından kolayca anlaşılır olabilmesi, sınama verisi veya yeni veriler üzerinde belli oranda geçerli olması, yeni, yararlı ve kullanılabilir olmasına bağlı kılınmıştır.

4.2. Verilerin Hazırlanması

Veri madenciliğinin en önemli aşamalarından bir tanesi olan verinin hazırlanması aşamasında kurumların mevcut bilgi sistemlerinin ürettiği sayısal bilgilerin analizi yapılarak mevcut problem ile arasında ilişki kurulmuştur. Bu aşama toplama, birleştirme ve temizleme, dönüştürme adımlarını içermektedir.

4.2.1.Toplama

Tez çalışmasında öncelikle uzun yıllar boyunca Malatya da yapılan meteoroloji uygulamalarında elde edilen verilerin saklandığı 1. veri tabanı oluşturulmuştur. Bu veri tabanı meteoroloji uygulamalarında elde edilen ortalama sıcaklık(Tablo 4.1), ortalama maksimum sıcaklık(Tablo 4.2), ortalama minimum sıcaklık(Tablo 4.3), nisbi nem(Tablo 4.4), güneşlenme süresi(Tablo 4.5), yağış ortalaması (Tablo 4.6) ve rüzgar hızı(Tablo 4.7) verilerini içermektedir.

Tablo 4.1. Malatya ili ortalama sıcaklık değerleri[13],[14]

	ORTALAMA SICAKLIK (°C)AYLAR											
YILLAR	1	2	3	4	5	6	7	8	9	10	11	12
1971	1,6	1,9	7,3	11,2	18,7	22,0	27,7	24,9	24,1	12,6	7,3	0,0
1972	-7,0	-4,0	6,6	12,0	16,2	21,2	27,8	26,7	22,7	17,0	5,7	1,4
1973	-3,3	4,3	6,4	11,9	19,2	21,5	26,6	27,1	22,8	16,0	3,3	-0,4
1974	-5,3	-2,5	8,2	10,6	19,6	25,1	26,6	25,5	19,6	18,8	7,8	3,3
1975	-1,2	0,0	7,7	15,1	16,8	23,8	28,0	26,5	21,4	14,4	7,3	-0,2
1976	-1,5	-4,0	4,4	11,0	16,5	22,1	24,6	25,4	20,0	14,6	8,8	2,7
1977	-1,6	5,5	6,9	13,0	17,7	22,9	25,8	26,4	22,3	12,1	9,0	0,7
1978	0,8	5,2	8,4	12,0	18,7	21,6	27,9	24,5	21,6	16,1	5,6	3,7
1979	2,0	5,3	9,0	13,4	18,0	22,5	25,2	28,0	23,7	15,1	8,8	2,1
1980	-2,3	0,2	5,4	11,5	17,7	24,0	28,9	26,4	21,2	15,1	8,7	2,5
1981	2,7	4,2	8,4	11,8	14,9	22,8	27,7	26,7	23,2	16,5	6,1	6,4
1982	1,3	-0,1	5,1	13,7	17,1	22,6	25,4	25,4	22,4	15,0	5,7	1,6
1983	-3,0	0,4	6,2	13,2	16,9	21,7	26,7	25,4	21,9	14,0	9,3	2,1
1984	2,7	4,2	8,0	11,6	17,5	25,0	27,5	23,9	24,0	15,3	7,4	-2,0
1985	1,5	-1,7	3,1	14,7	20,5	24,3	25,3	28,9	22,4	13,1	10,2	2,6
1986	1,5	3,3	8,4	15,1	14,0	22,1	28,7	29,0	23,9	13,3	5,5	2,2
1987	0,3	3,9	3,2	11,4	18,5	23,5	27,7	26,1	21,7	13,7	5,8	2,2
1988	-0,8	1,9	4,7	11,4	17,5	22,0	26,5	26,3	21,5	13,8	2,8	2,2
1989	-1,2	2,6	9,6	18,1	20,0	24,4	28,3	28,1	22,3	12,7	7,6	2,2
1990	-2,7	1,9	7,9	12,2	18,1	22,7	29,0	26,0	21,7	14,8	7,5	2,2
1991	-0,2	-1,2	7,0	13,3	17,2	24,0	27,8	27,5	21,8	16,3	10,6	2,2
1992	-2,8	-5,0	3,3	12,5	16,1	21,2	24,9	26,4	20,8	16,8	5,0	2,2
1993	-3,1	-1,4	5,2	12,3	15,4	21,7	27,2	26,0	22,5	16,6	4,4	2,2
1994	3,6	2,1	8,4	16,6	19,5	23,4	27,3	25,5	24,9	17,5	6,5	2,2
1995	1,8	4,3	8,6	11,8	18,9	23,0	25,8	27,3	22,3	14,4	5,7	2,2
1996	2,1	3,7	6,7	10,7	19,9	22,4	28,7	27,2	21,3	14,0	8,3	2,2
1997	2,8	0,1	3,2	9,8	19,7	22,5	26,2	26,3	19,8	15,2	8,0	2,2
1998	0,8	1,7	6,0	14,0	17,2	24,5	28,8	28,6	22,8	16,6	11,3	2,2
1999	2,7	4,0	7,5	13,5	19,3	23,1	27,7	27,7	21,7	16,2	8,3	2,2
2000	-1,4	-0,8	4,7	13,8	17,8	23,5	30,2	26,1	22,4	14,3	9,0	2,2
2001	3,0	4,0	12,0	14,5	16,0	24,9	28,8	28,4	23,9	16,0	6,1	2,2
2002	-2,3	5,1	9,3	12,1	18,1	23,7	28,2	26,5	23,1	16,3	9,7	2,2
2003	3,0	0,4	2,8	12,2	20,3	23,9	27,1	28,2	21,7	17,2	8,1	2,2
2004	1,2	1,8	8,9	12,4	17,3	23,6	27,6	27,5	22,9	17,5	7,1	2,2
2005	0,8	2,1	7,4	14,0	18,4	22,9	29,1	28,6	22,4	13,7	6,3	2,2
2006	-1,0	3,2	9,1	14,2	19,0	26,9	27,2	37,0	22,4	16,8	7,4	2,2

Tablo 4.2. Malatya ili ortalama maksimum sıcaklık değerleri[13],[14]

	ORTALAMA MAX SICAKLIK (°C)AYLAR											
YILLAR	1	2	3	4	5	6	7	8	9	10	11	12
1971	5,6	5,7	12,4	15,6	24,3	28,3	34,0	31,2	30,4	18,6	11,9	3,1
1972	-3,2	-0,6	11,7	19,0	21,0	26,5	33,5	32,3	28,7	28,7	10,7	6,5
1973	0,2	8,4	11,0	17,0	25,0	27,4	32,9	33,3	29,2	21,7	7,4	2,9
1974	-2,0	0,5	11,9	15,1	25,1	30,9	33,0	31,6	25,8	24,8	11,7	6,2
1975	2,2	3,2	12,9	20,4	21,6	29,4	33,9	33,3	27,6	20,7	12,2	3,4
1976	1,3	-0,8	8,9	15,2	21,6	27,7	30,9	31,9	26,4	20,0	14,3	5,8
1977	1,5	9,3	11,3	18,3	22,9	28,9	32,3	33,6	28,8	18,9	14,8	3,2
1978	3,1	9,1	12,9	17,2	24,5	27,7	34,6	31,6	28,9	22,4	11,8	6,9
1979	5,3	9,0	13,8	18,9	23,9	28,3	31,7	35,0	31,0	20,6	13,2	5,3
1980	0,6	3,5	9,7	16,5	23,1	30,1	35,9	33,5	28,1	21,6	15,4	5,5
1981	5,6	8,0	12,7	17,4	20,1	28,9	34,5	33,4	30,4	23,0	11,2	9,8
1982	5,2	4,2	10,4	18,6	22,9	29,0	32,0	32,5	29,2	21,4	11,0	6,5
1983	0,6	4,1	10,6	18,6	23,0	27,6	33,5	32,4	29,2	20,6	12,8	5,0
1984	6,2	8,9	12,5	17,0	23,6	29,8	34,9	37,1	31,7	22,8	11,4	1,5
1985	4,9	2,4	7,9	20,9	26,6	31,4	33,3	37,1	30,6	20,0	15,8	5,8
1986	5,3	7,4	14,3	21,8	19,6	28,7	35,9	36,7	31,5	19,8	10,7	6,9
1987	3,3	8,3	8,2	17,0	25,2	30,4	34,9	33,5	29,7	19,6	11,2	4,9
1988	2,3	5,8	8,9	16,9	24,3	28,6	33,6	33,8	29,6	19,6	6,9	7,1
1989	3,6	8,0	14,7	24,7	27,0	31,7	36,3	36,6	30,7	19,5	18,8	3,8
1990	1,3	7,0	14,9	18,1	24,2	39,2	36,8	34,8	30,2	22,4	12,8	8,9
1991	3,0	3,1	11,4	18,0	23,5	30,4	34,2	23,6	28,4	21,6	10,6	3,3
1992	0,3	-1,0	8,3	18,6	21,5	26,5	30,7	32,2	26,8	23,1	9,6	0,6
1993	0,6	2,6	10,3	17,8	20,5	27,7	33,6	32,2	28,9	23,0	8,4	7,6
1994	7,2	5,7	13,3	22,0	25,5	29,5	33,4	32,0	31,3	22,8	10,4	1,8
1995	4,6	9,0	13,8	16,8	24,7	29,2	31,9	33,5	28,6	20,3	10,1	5,8
1996	5,1	7,5	10,4	15,1	25,8	28,6	35,3	33,2	27,6	19,6	13,1	14,4
1997	7,4	4,5	8,7	15,4	25,9	29,1	32,3	32,6	26,2	20,1	12,4	6,3
1998	4,3	7,0	11,6	19,6	22,8	31,0	35,6	35,3	29,5	23,7	16,7	8,4
1999	6,2	9,3	13,1	19,5	26,4	29,7	34,2	34,5	28,8	22,5	14,4	7,4
2000	2,9	3,3	10,7	19,6	24,2	30,3	37,6	33,4	29,2	19,9	15,9	5,7
2001	7,0	9,2	17,2	20,0	22,0	31,6	35,7	35,4	31,1	22,8	10,7	6,4
2002	1,2	10,4	15,2	17,3	23,9	29,6	34,3	32,7	29,4	22,8	15,8	2,2
2003	5,8	4,2	7,1	17,6	26,3	30,1	33,4	34,7	28,7	23,4	12,9	6,1
2004	4,0	5,5	14,7	18,4	23,3	29,7	33,9	33,9	30,1	23,8	11,1	4,5
2005	4,6	6,1	12,1	19,8	24,4	28,9	35,5	35,2	28,9	20,1	10,4	4,5
2006	2,1	7,4	14,5	19,6	25,3	33,5	33,5	37,7	29,2	22,5	13,1	7,8

Tablo 4.3. Malatya ili ortalama minimum sıcaklık değerleri[13],[14]

	ORTALAMA MİN.SICAKLIK. (°C) AYLAR											
YILLAR	1	2	3	4	5	6	7	8	9	10	11	12
1971	-2,0	-1,5	3,2	6,6	12,4	15,4	20,3	18,6	17,0	7,2	3,6	-2,5
1972	-10,4	-7,4	1,7	8,8	10,7	15,3	21,0	19,6	16,0	11,3	1,5	-2,8
1973	-6,4	0,2	1,7	6,7	12,5	15,0	19,7	19,5	15,6	10,5	-0,2	-3,3
1974	-7,7	-5,3	4,5	6,2	12,5	17,6	19,1	18,8	13,4	12,5	4,3	0,6
1975	-4,6	-2,8	2,4	9,6	10,8	16,3	20,8	19,2	14,8	8,0	3,0	-3,3
1976	-3,8	-7,0	0,3	6,9	11,0	15,8	18,0	18,6	13,9	10,2	4,6	-0,2
1977	-4,5	1,8	2,9	7,9	12,0	16,5	18,8	19,3	13,5	6,4	4,1	-1,8
1978	-1,8	1,5	4,0	7,0	11,9	14,8	20,5	17,9	14,5	10,7	0,5	0,7
1979	-1,1	1,9	4,0	7,8	11,9	15,6	18,4	20,8	16,4	10,5	5,5	-0,9
1980	-5,3	-3,2	1,5	6,9	11,6	17,1	21,9	19,5	14,4	9,8	4,7	-0,1
1981	-0,1	0,8	4,9	6,6	9,6	16,5	20,6	20,4	16,4	11,1	2,3	3,6
1982	-1,7	-3,8	0,5	9,2	11,7	16,1	18,9	18,7	15,8	9,8	1,9	-1,8
1983	-5,0	-2,6	2,1	8,2	11,6	15,1	19,8	18,6	15,1	8,8	6,4	-0,2
1984	0,1	0,5	4,4	7,0	11,7	15,9	20,1	17,2	16,3	9,0	4,3	-4,6
1985	-1,4	-4,9	-1,4	8,9	13,9	17,0	17,6	20,8	15,1	7,9	6,0	-0,1
1986	-1,3	0,3	3,0	9,1	9,2	15,7	21,3	21,8	17,5	8,8	1,9	-1,1
1987	-2,0	0,6	-0,8	6,1	12,1	16,4	20,1	19,2	14,5	8,9	2,1	-0,3
1988	-2,9	-1,0	1,0	6,4	12,0	16,3	19,8	19,8	15,1	9,5	1,5	1,3
1989	-4,4	-1,8	5,3	11,2	13,2	16,4	20,2	20,1	15,8	7,9	3,8	-1,7
1990	-5,4	-2,0	2,4	6,6	10,6	15,3	20,5	12,1	14,1	9,1	4,0	9,5
1991	-2,8	-4,8	2,5	7,8	10,3	16,5	20,2	20,2	15,2	11,3	4,3	-1,1
1992	-5,5	-8,1	-1,0	6,5	10,1	10,9	17,9	19,6	14,7	10,9	1,5	-3,6
1993	-6,3	-4,8	0,5	7,0	10,3	15,0	19,2	19,1	15,6	10,4	1,4	1,9
1994	0,5	-1,1	3,4	10,3	13,1	15,7	19,9	18,4	18,1	12,3	1,3	-4,8
1995	-0,7	0,3	3,6	6,9	12,5	16,7	18,7	20,2	15,8	9,0	2,5	-1,3
1996	-0,2	0,3	3,3	6,0	13,2	14,6	21,2	19,9	14,6	9,3	4,4	3,6
1997	-0,7	-3,9	-1,5	4,3	12,5	15,3	18,8	19,0	12,7	10,5	4,0	0,3
1998	-2,4	-3,2	1,3	8,0	11,6	17,0	20,7	20,7	15,5	9,9	6,6	1,5
1999	-0,3	-0,6	2,2	7,1	11,5	16,0	20,2	20,2	14,5	10,2	3,3	-0,3
2000	-4,9	-4,0	-0,7	7,9	11,0	15,4	21,9	18,9	15,6	9,2	3,7	0,3
2001	-0,3	-0,7	6,5	8,7	10,5	16,9	20,8	20,8	16,1	9,8	2,1	0,9
2002	-5,6	0,4	3,9	7,1	11,6	20,6	20,6	19,8	16,1	10,8	4,6	-5,5
2003	0,3	-2,8	-0,9	7,1	13,6	16,6	19,9	20,9	15,4	11,8	4,4	0,0
2004	-1,3	-1,7	3,6	6,4	11,8	16,7	20,3	17,0	21,8	11,8	4,0	-2,6
2005	-2,3	-1,5	3,1	8,6	12,1	16,2	21,6	21,9	16,3	8,6	2,8	-0,1
2006	-3,6	-0,6	4,4	9,2	12,4	19,3	20,4	23,3	15,7	11,9	3,1	-2,8

Tablo 4.4. Malatya ili nisbi nem deęerleri[13],[14]

	ORTALAMA NİSBİ NEM (%)AYLAR											
YILLAR	1	2	3	4	5	6	7	8	9	10	11	12
1971	72,4	68,3	61,6	64,7	46,4	37,1	31,3	38,9	31,1	53,0	66,5	75,1
1972	77,2	77,6	56,7	57,5	59,4	54,3	31,2	33,3	37,9	48,2	71,2	57,9
1973	75,1	69,1	59,4	57,3	39,4	39,6	29,7	26,3	30,3	48,8	71,9	80,3
1974	83,6	84,9	68,1	57,2	42,8	32,4	28,4	34,5	40,7	42,9	75,0	77,2
1975	73,9	76,6	52,0	51,2	58,4	39,5	33,0	31,2	38,6	44,0	62,9	72,6
1976	81,5	77,6	69,8	70,8	57,9	40,0	35,1	37,5	45,7	61,9	72,4	78,8
1977	73,1	67,7	60,7	52,6	54,1	39,9	35,8	34,7	44,1	47,2	59,9	79,4
1978	83,4	72,1	66,6	56,6	41,1	35,9	38,8	32,9	40,2	53,5	52,2	79,4
1979	74,3	74,5	59,6	50,0	55,5	37,2	35,2	29,9	33,7	62,0	82,7	78,7
1980	75,1	76,5	69,8	60,8	51,3	33,8	28,4	35,0	39,1	46,2	70,2	74,7
1981	80,2	72,0	65,0	49,1	54,4	42,6	31,6	34,1	33,5	54,4	67,2	74,2
1982	66,6	62,2	56,7	60,0	56,8	40,1	38,6	35,1	40,1	48,6	58,4	67,1
1983	67,0	71,7	61,1	49,1	58,6	45,9	34,6	34,4	33,6	55,6	75,7	78,7
1984	75,2	60,7	66,7	59,5	46,1	37,0	33,2	35,6	33,8	48,5	76,0	75,6
1985	82,5	73,1	70,3	53,8	53,3	44,8	41,0	31,7	34,2	61,3	76,0	74,4
1986	78,7	78,7	58,0	48,1	63,7	47,3	52,2	38,0	49,1	71,3	71,9	73,0
1987	76,9	70,8	62,5	54,3	57,0	45,3	41,2	39,1	39,6	61,1	75,3	73,8
1988	77,8	72,3	67,7	61,6	56,6	51,6	50,4	47,3	45,0	65,1	72,0	75,2
1989	56,3	53,6	59,8	40,5	40,0	38,1	56,0	37,5	45,2	63,5	71,8	78,9
1990	69,4	67,4	50,4	53,9	44,0	36,1	31,3	31,8	34,8	49,9	69,2	63,3
1991	74,4	77,1	73,5	64,1	51,4	41,7	37,7	32,0	38,1	58,0	78,8	75,9
1992	66,1	69,5	59,5	42,9	49,3	47,2	36,5	35,2	40,8	50,5	67,7	77,7
1993	71,5	70,0	56,1	52,8	66,0	44,9	37,1	40,7	38,8	44,0	68,7	76,7
1994	72,8	64,3	54,3	48,1	43,5	34,2	35,1	33,6	33,2	53,3	69,4	76,7
1995	75,0	69,4	56,9	57,9	52,9	43,1	40,7	31,5	38,4	53,0	69,5	69,4
1996	76,5	71,1	71,2	60,6	49,9	38,3	31,6	32,4	41,8	59,9	70,2	80,1
1997	65,3	67,1	55,9	57,4	45,0	41,2	30,5	33,5	42,0	63,4	70,1	74,4
1998	67,7	57,4	62,2	58,0	61,6	42,4	30,6	28,4	40,3	47,2	65,7	76,0
1999	74,0	61,8	54,5	50,3	39,7	39,6	35,3	32,2	35,1	47,3	49,5	70,9
2000	67,5	64,9	53,7	53,8	43,0	33,0	19,8	30,1	33,6	54,3	54,5	67,8
2001	65,1	62,3	52,9	49,3	53,3	28,1	24,9	26,6	27,4	38,8	60,9	79,4
2002	75,4	65,2	58,9	61,4	51,0	44,4	35,0	37,5	48,8	57,7	65,6	74,0
2003	78,1	70,3	70,8	58,8	45,1	36,0	35,2	40,3	45,6	55,6	72,5	75,4
2004	79,0	71,6	51,1	50,0	57,4	41,5	29,4	39,2	41,3	57,5	73,6	69,9
2005	72,9	64,8	59,6	50,1	50,1	40,1	28,4	39,4	44,4	53,4	72,4	83,2
2006	71,7	67,7	77,8	62,9	49,6	33,9	38,1	28,5	42,3	63,2	62,5	54,4

Tablo 4.5. Malatya ili güneşlenme süresi değerleri [13],[14]

	ORTALAMA GÜNEŞLENME SÜRESİ AYLAR											
YILLAR	1	2	3	4	5	6	7	8	9	10	11	12
1986	2,7	3,2	7,3	8,6	7,1	11,2	13,2	11,9	8,9	6,3	5,4	4,4
1987	1,2	4,1	5,2	6,9	10,7	12,1	12,5	10,7	9,5	5,8	5,6	1,7
1988	2,1	2,8	4,2	6,4	9,3	9,9	12,0	11,4	9,4	6,0	2,9	2,5
1989	5,9	6,5	5,9	9,0	11,5	11,7	13,2	11,9	9,4	6,8	4,5	2,1
1990	3,4	5,4	7,5	6,6	9,7	11,8	12,4	11,6	9,8	6,5	4,2	4,0
1991	4,2	4,4	4,6	6,7	9,0	11,6	12,0	11,7	10,5	6,3	3,5	1,9
1992	4,1	3,0	5,8	8,6	7,7	9,7	12,5	11,7	9,8	8,2	4,0	0,8
1993	4,0	3,7	6,6	7,1	7,3	11,3	12,7	11,5	10,6	9,0	4,5	3,0
1994	4,4	4,9	6,6	8,6	10,3	12,5	12,4	12,1	9,5	7,4	4,5	1,8
1995	2,6	5,7	7,3	7,3	9,8	10,8	12,3	12,0	10,2	8,4	5,3	4,1
1996	2,6	4,3	3,6	6,1	10,6	8,1	12,8	12,1	9,6	6,4	5,6	2,2
1997	5,7	5,5	6,2	6,7	9,9	10,5	12,5	12,1	10,0	6,3	4,8	3,0
1998	4,0	6,6	5,9	6,9	6,6	11,3	11,7	12,1	10,1	8,8	5,4	3,3
1999	2,6	5,4	6,3	8,1	11,0	10,7	11,6	12,0	10,5	7,8	5,7	3,9
2000	3,6	5,2	7,2	6,6	9,9	11,9	12,4	11,5	10,0	6,9	7,3	1,8
2001	3,8	5,4	6,3	7,8	8,3	11,7	11,9	11,8	10,5	8,5	5,1	2,1
2002	4,4	6,4	6,2	6,7	10,4	11,7	12,2	11,4	10,3	7,9	6,4	4,0
2003	2,5	3,5	4,6	5,9	5,4	12,5	12,9	12,3	10,6	7,4	4,9	3,1
2004	1 ,6	2,7	7,1	8,6	14,2	12,1	14,5	12,1	11,1	8,1	4,3	9,5
2005	3,9	4,2	5,6	7,5	14,2	11,5	12,4	11,6	10,4	7,8	3,9	1,2
2006	3,3	4,5	6,2	6,1	10,4	12,0	11,8	11,5	10,4	6,8	6,9	7,2

Tablo 4.6. Malatya ili yağış değerleri[13],[14]

	TOPLAM YAĞIŞ (mm)AYLAR											
YILLAR	1	2	3	4	5	6	7	8	9	10	11	12
1971	7,2	25,0	69,1	137,4	34,0	3,6	-	23,6	0,7	33,4	43,3	58,2
1972	24,8	17,9	21,5	35,3	89,8	46,2	0,2	-	26,8	16,8	38,0	3,5
1973	35,8	9,3	47,5	93,5	16,5	28,3	0,4	-	1,2	45,4	65,3	37,0
1974	49,8	15,8	60,2	100,1	26,0	2,3	0,0	0,6	9,1	18,5	25,1	37,3
1975	18,3	44,8	36,7	72,5	59,5	9,0	0,0	0,0	2,0	1,3	28,7	24,9
1976	51,0	27,5	38,0	141,2	95,5	7,5	5,6	-	13,1	97,5	33,6	47,2
1977	42,5	17,9	91,2	54,0	46,8	1,7	-	-	0,7	11,3	4,7	53,0
1978	68,7	46,7	68,1	67,3	19,8	21,1	-	-	5,7	29,6	0,4	22,8
1979	70,9	39,7	44,8	49,2	69,8	35,2	4,0	0,0	0,0	67,1	122,3	37,2
1980	35,4	26,7	121,7	51,6	46,1	5,0	1,6	0,5	-	8,4	52,4	77,9
1981	33,6	52,9	78,4	56,5	39,0	27,2	8,2	-	1,4	56,4	21,2	45,6
1982	26,7	14,3	57,1	71,7	62,6	18,4	12,0	0,2	3,5	2,6	29,9	10,3
1983	21,2	39,9	80,4	60,2	130,9	35,3	0,1	-	3,0	52,4	42,8	35,8
1984	75,7	14,2	37,7	79,4	42,8	41,8	0,2	0,4	-	9,1	34,4	54,6
1985	10,4	41,9	36,2	29,4	39,3	7,3	-	0,0	-	26,0	37,6	27,5
1986	26,3	66,7	13,4	2,9	70,2	62,7	-	2,6	5,0	77,2	64,2	34,6
1987	31,7	41,7	57,6	53,6	19,5	3,2	4,3	-	0,6	57,2	58,4	92,9
1988	33,6	39,9	92,1	120,1	40,7	40,8	5,3	0,1	2,4	129,5	53,7	39,2
1989	3,7	15,1	64,2	12,6	5,1	25,1	-	0,3	1,3	66,5	60,0	23,4
1990	23,5	57,2	11,8	50,8	28,0	22,3	-	0,0	1,7	16,3	52,7	15,5
1991	28,8	33,3	60,8	44,4	12,0	23,5	-	-	23,7	12,1	38,8	67,3
1992	13,3	49,0	9,4	11,0	76,7	59,1	-	-	17,2	23,5	90,6	67,3
1993	31,8	59,3	43,9	19,7	158,2	30,7	-	2,2	-	-	32,3	18,8
1994	40,7	48,3	14,0	44,6	19,8	1,3	3,2	1,7	2,2	25,1	75,1	60,8
1995	42,2	32,7	25,0	71,7	61,0	38,6	2,5	1,9	3,0	35,2	65,4	0,8
1996	47,1	34,4	95,5	69,5	39,1	0,6	1,4	1,0	9,0	60,9	17,8	62,2
1997	18,0	53,2	41,2	44,3	41,8	18,5	-	-	22,6	67,6	24,2	49,3
1998	34,9	9,8	66,9	83,9	86,1	12,7	0,9	-	2,6	10,0	56,2	43,9
1999	43,6	12,7	46,9	54,1	5,5	18,7	5,8	0,9	3,5	24,8	9,6	27,3
2000	71,1	29,9	34,8	32,7	28,6	1,4	-	12,8	9,7	67,8	5,9	38,0
2001	4,2	55,8	49,5	41,5	60,8	-	7,6	-	2,0	16,2	19,4	59,2
2002	41,7	33,4	57,9	82,1	39,1	2,9	8,9	1,8	10,2	11,2	10,4	51,7
2003	25,1	79,0	103,6	64,3	6,7	7,8	3,2	-	13,6	38,0	33,5	20,2
2004	73,8	49,8	3,0	84,5	37,1	1,3	0,5	-	-	15,7	86,4	18,2
2005	33,1	14,1	40,4	46,0	60,5	5,8	-	3,3	4,3	59,3	31,7	23,0
2006	32,3	53,3	29,6	67,0	26,1	8,5	7,2	0,2	6,7	84,7	38,0	0,3

Tablo 4.7. Malatya ili rüzgar hızı değerleri[13],[14]

ORTALAMA RÜZGARIN HIZI												
YILLAR	1	2	3	4	5	6	7	8	9	10	11	12
1991	1,0	1,5	2,1	2,2	2,9	2,5	2,5	2,3	2,2	1,7	1,1	1,5
1992	2,0	1,7	2,1	2,7	2,1	2,0	2,7	2,3	2,1	1,8	1,7	1,2
1993	1,5	1,8	2,3	2,9	1,9	2,1	2,4	2,1	2,0	1,5	1,6	1,1
1994	1,3	1,6	2,1	2,7	2,1	3,0	2,3	2,4	2,0	1,5	1,5	1,3
1995	1,3	1,5	2,4	2,2	1,9	2,3	2,4	2,1	2,1	1,8	2,1	1,3
1996	1,3	1,5	1,8	2,1	2,1	2,5	2,2	2,0	2,0	1,5	0,9	1,4
1997	1,6	1,5	2,2	2,6	2,0	2,2	2,5	2,0	2,0	1,4	1,1	1,2
1998	1,5	1,8	2,2	1,8	1,6	1,9	2,0	1,7	1,9	1,6	1,3	1,2
1999	1,2	1,9	2,0	1,8	2,2	2,1	1,9	1,8	1,7	1,3	1,5	1,1
2000	1,5	1,5	1,9	2,1	2,0	2,3	2,0	2,0	1,7	1,3	1,1	1,1
2001	1,0	1,6	1,7	2,2	1,9	2,2	1,9	2,0	1,7	1,4	1,4	1,2
2002	1,5	1,4	2,0	2,0	1,9	2,4	2,0	2,0	1,6	1,4	1,2	1,3
2003	1,2	1,6	1,6	1,9	1,9	1,9	2,1	1,7	1,6	1,3	1,0	0,9
2004	1,1	1,5	1,8	1,9	1,7	1,9	1,8	1,6	1,9	1,2	1,2	1,2
2005	1,4	1,6	1,9	1,8	1,8	1,9	1,8	1,6	1,6	1,4	0,9	1,0
2006	1,0	1,4	1,7	1,7	1,3	1,8	1,8	1,6	1,7	1,1	1,4	1,3

Daha sonra tablo 4.8’de tarımsal ürünlerin yetişmesi için gerekli olan iklimsel özelliklerin saklandığı 2. veri tabanı oluşturulmuştur.

Tablo 4.8. Tarım ürünlerinin iklim istekleri değerleri[11],[12], [16],[17],[18]

ÜRÜN	Ort Sıcaklık		Min Sıcaklık	Max Sıcaklık	Nisbi Nem %		Güneşlenme Süresi		Rüzgar Hızı Km/Saat	Toplam Yağış		Toplam Sıcaklık	
	Min	Max	°C	°C	Min	Max	Min	Max		Min	Max	Min	Max
Pırasa	15	20	10	30	60	70	3	9	4	500	700	1750	2250
Pamuk	15	25	5	34	50	70	4	11	4	400	700	2300	5000
Portakal	16	26	6	38	65	80	3	11	3	500	800	2300	5000
Buğday	10	15	5	40	50	70	3	10	3	300	600	1750	2250
Soya Fasulyesi	15	25	12	37	60	80	3	10	4	500	700	1750	2250

Veri tabanında tarımsal ürünlerin yetişmesi için gerekli olan optimum (ortalama) sıcaklık, gerekli en düşük (min) sıcaklık, dayanabileceği en yüksek (max) sıcaklık, gerekli nem değeri aralığı, gerekli güneşlenme süresi, yıl içerisindeki toplam yağış miktarı ve yıl içerisindeki toplam sıcaklık miktarı parametrelerini içermektedir(Tablo 4.8).

4.2.2.Birleştirme ve Temizleme

Birleştirme ve temizleme adımının ilk aşaması olan birleştirme aşamasında farklı kaynaklardan alınarak toplanan verilerde bulunan farklılıklar giderilmiştir. Bu farklılıklar aynı değeri ifade eden başlıkların eşleşmesi anlamını taşımaktadır(Optimum sıcaklık=Ortalama sıcaklık, En düşük sıcaklık=Minimum sıcaklık gibi...).

Tablo 4.9’da temizleme aşamasında ürünlerin verimli olarak yetişmesini sağlayan etkin ay aralıkları için 1 parametresi, ürünün yetişmesine etki etmeyen ay aralıkları için 0 parametresi tanımlanmıştır.

Tablo 4.9.Ürünlerin yetişmesini sağlayan etkin ay aralıkları[16],[17],[18]

Ürün/Ay	Ocak	Şubat	Mart	Nisan	Mayıs	Haziran	Temmuz	Ağustos	Eylül	Ekim	Kasım	Aralık
Buğday	0	0	1	1	1	1	1	0	0	0	0	0
Domates	0	0	0	1	1	1	1	1	1	0	0	0
Kayısı	0	0	1	1	1	1	1	1	1	1	0	0
Üzüm	0	0	0	1	1	1	1	1	1	0	0	0
Elma	0	0	1	1	1	1	1	1	1	0	0	0

4.2.3. Dönüştürme

Dönüştürme adımı tablo 4.10’da verinin yapısıyla oynanmadan uygulanacak olan kümeleme, sınıflandırma ve ilişkilendirme modelleri için gösterimi belirlenmiştir.

Tablo 4.10. Meteorolojik verilerin dönüştürme aşamasındaki gösterimi[11],[12], [13],[14]

AYLAR	Ortalama Sıcaklık °C	Min Sıcaklık °C	Max Sıcaklık °C	Nisbi Nem %	Güneşlenme Süresi	Yağış Toplamı m ² /kg	Rüzgar Hızı Km/saat
OCAK	-0,1	-3	3,3	74	3,5	35,3	1,3
ŞUBAT	1,6	-2	5,7	69	4,6	36,2	1,6
MART	6,8	2,3	11,8	61	6	51,4	2
NİSAN	12,9	7,6	18,3	55	7,2	61,1	2,2
MAYIS	17,9	11,7	23,8	51	9,6	48,4	2
HAZİRAN	23,1	16,1	29,5	40	11,2	19,3	2,2
TEMMUZ	27,3	20	34	35	12,4	3,6	2,1
AĞUSTOS	27	19,5	33,4	34	11,7	2,6	2
EYLÜL	22,3	15,5	29,2	38	10	6,7	1,9
EKİM	15,3	9,9	21,6	54	7,3	39,3	1,5
KASIM	7,3	3,3	12,3	69	4,9	41,8	1,3
ARALIK	2	-0,6	5,6	74	3,2	38,5	1,2

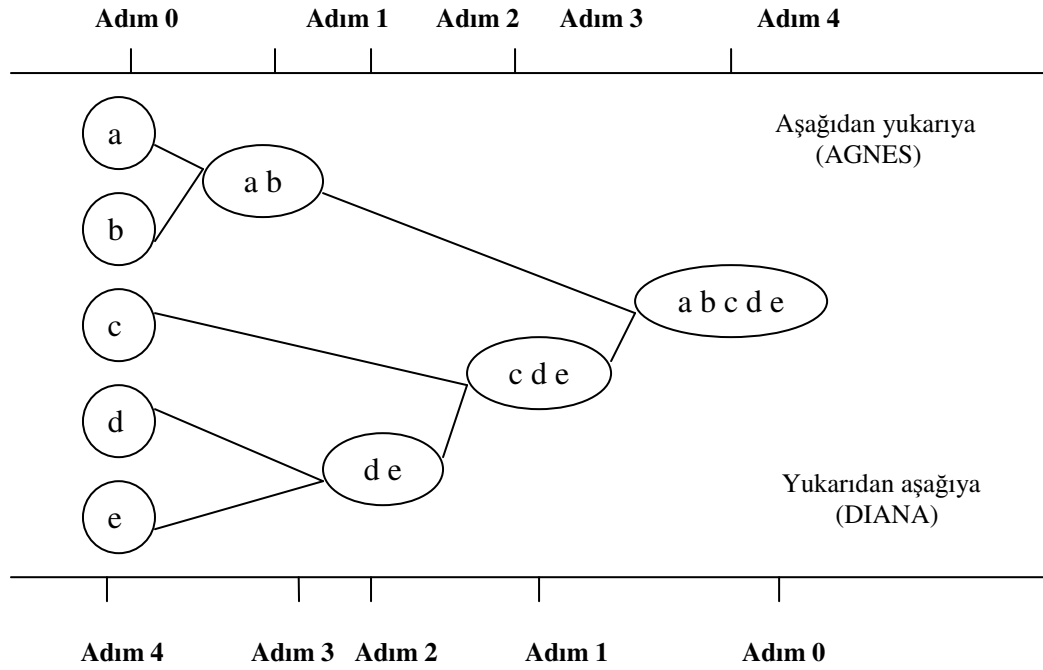
4.3. Modelin kurulması ve değerlendirilmesi

Dönüştürme aşamasından gelen veri üzerinde uygulanacak en uygun modelin seçilmesi adımını kapsamaktadır. Uygulamanın bu bölümünde kümeleme ve sınıflandırma modelleri kurulmuş ve değerlendirilmiştir.

Kümeleştirmenin amacı kayıtları ortak özellikleri olanlar bir arada olacak şekilde gruplandırmaktır.[15] Yani kayıtların en sık görülen ortak özelliklerinin belirlenmesi ve bu ortak özelliklere göre kayıt gruplarının oluşturulmasıdır.

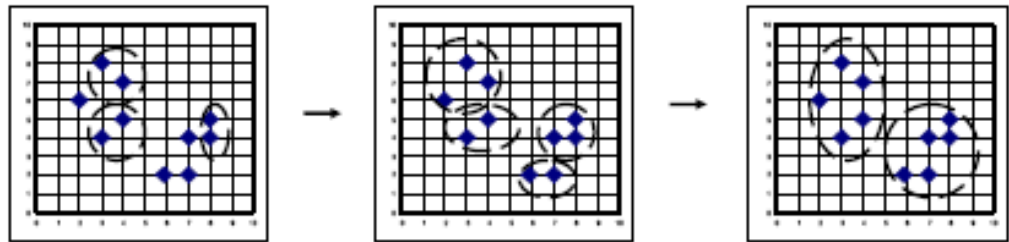
Tezde kümeleme işlemi, ürünlerin en sık görülen ortak özelliklerine göre bizim daha önceden belirlediğimiz tahıllar, baklagiller, endüstriyel bitkiler, meyvesi yenenler, yumrusu yenenler, yaprağı yenenler, yumuşak çekirdekli, sert çekirdekli ve üzüm bitkileri adlı 9 gruba ayrılmıştır. Temel kümeleme yöntemlerini daha önce belirttiğimiz bölümler, hiyerarşik yöntemler, yoğunluk tabanlı yöntemler ve model tabanlı yöntemler olmak üzere dört grupta inceleyebiliriz.

Tezde kümeleme işlemi için Hiyerarşik kümeleme yöntemi kullanılmıştır. Hiyerarşik kümeleme yönteminde Şekil 4.1’de veri kümeleri (yada nesneler) önceden belirlenmiş bir kritere göre hiyerarşik olarak ayrılır.



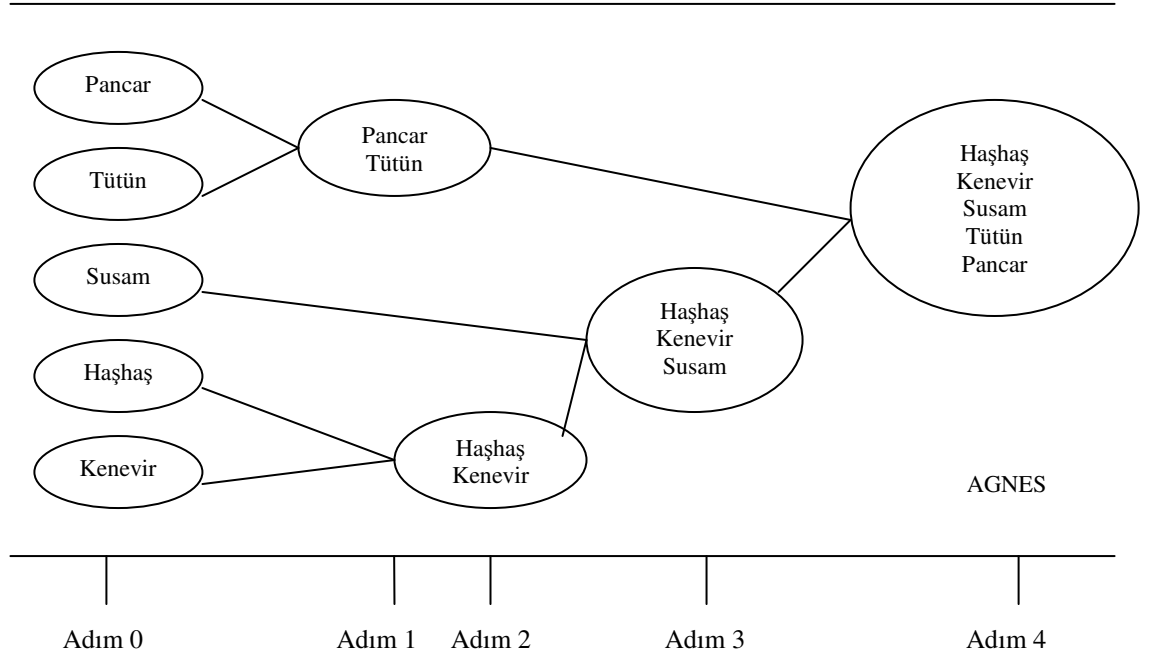
Şekil 4.1. Hiyerarşik kümeleme yöntemi

Şekil 4.2’de AGNES yönteminde, birinci adımda her nesne bir demet oluşturur. Aralarında en az uzaklık bulunan kümeler her adımda birleştirilir. Bütün nesneler tek bir küme içinde kalana kadar yada istenilen sayıda küme elde edilene kadar aşağıdan yukarıya doğru birleştirme işlemi devam eder.



Şekil 4.2. AGNES kümeleme yöntemi

80 tarım ürününün kümelenmesinde birinci adımda her tarım ürünü bir demet oluşturdu daha sonra aralarında en az uzaklık bulunan ürünler her adımda birleştirildi ve önceden belirlenen 9 kümeye ulaşınca kadar bu işlem devam etti.



Şekil 4.3. Tarım ürünleri Kümeleme(AGNES)

Kümeleme modelinde, oluşan kümelerin kalitesi seçilen benzerlik ölçütüne ve bu ölçütün gerçekleşmesine bağlıdır. Uygulamamızda her ürün içinde sık geçen terimler bulundu ve bulunan bu terimlerden yararlanılarak bir benzerlik ölçütü geliştirildi. Bu ölçüt ile ürünler yapısal özelliklerine göre tablo 4.11’de 9 grupta kümelendirildi.

Tablo 4.11. Tarım ürünlerinin 9 kümede gösterimi

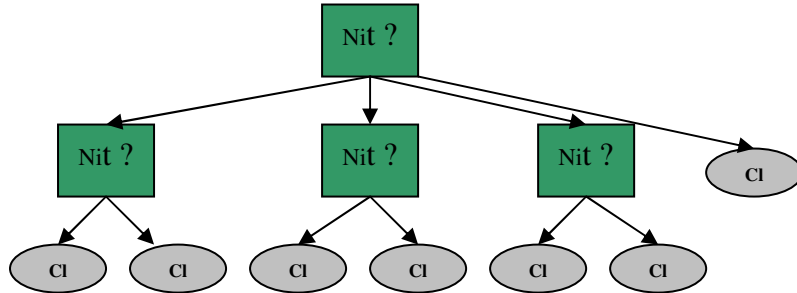
KÜMELEME MODELİ								
TAHILLAR	BAKLAGİLLER	ENDÜSTRİYEL BİTKİLER	MEYVESİ YENENLER	YUMRUSU YENİLENLER	YAPRAĞI YENİLENLER	YUMUŞAK ÇEKİRDEKLİLER	SERT ÇEKİRDEKLİLER	ÜZÜMSÜ BİTKİLER
Arpa	Bakla	Ay Çiçeği	Biber	Havuç	Dereotu	Armut	Badem	Çilek
Buğday	Bamya	Fiğ	Domates	Patates	Enginar	Ayva	Ceviz	Dut
Çavdar	Barbunya	Haşhaş	Hıyar	Sarımsak	İspanak	Elma	Erk	İncir
Çeltik	Bezelye	Kendir	Kabak	Soğan	Karnabahar	Limon	Fındık	Kivi
Dan	Börtlüce	Kenevir	Karpuz	Turp	Kereviz	Mandalina	Kayısı	Muz
Mısır	Fasulye	Keten	Kavun	Yer Fıstığı	Lahana	Muşmula	Kestane	Nar
Yulaf	Mercimek	Pamuk	Patlıcan		Marul	Portakal	Kızılçık	Üzüm
	Nohut	Soya Fasülyesi			Maydonoz	Trabzon Hurması	Kiraz	
		Şeker Pancarı			Nane	Turunç	Şeftali	
		Tütün			Pazı	Yeni Dünya	Vişne	
					Pırasa		Zerdali	
					Semiz Otu		Zeytin	

Sınıflandırma modeli, daha önce ayrılmış veriler üzerinde bir model kurar. Bu modeli daha önce yapılan ayrıştırmanın temellerini açıklamak veya yeni verileri sınıflandırmak amacı ile kullanabiliriz. Temel sınıflandırma modellerini karar ağaçları, yapay sinir ağları, bayes sınıflandırıcı ve bayes ağları olarak dört gruba ayırabiliriz.

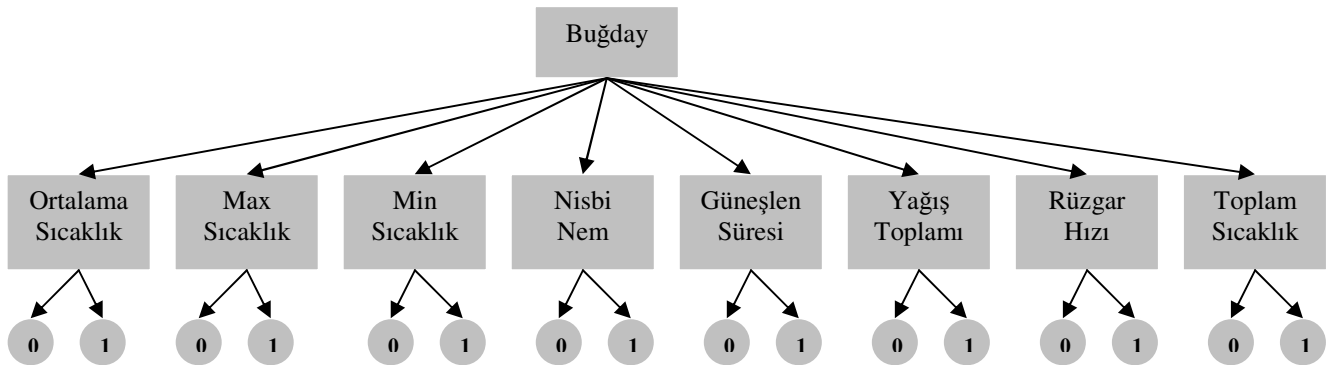
Uygulamada sınıflandırma yöntemlerinden karar ağaçları (decision trees) tekniği kullanıldı. Karar ağaçları algoritmaları verinin belirli sınıflara atanması için kullanılır.[7] Karar ağaçları tekniği ile istenilen örneğin nitelikleri ağaç boyunca sınanır ve ulaşılan yaprağın etiketi sınıf bilgisini verir.

Akış diyagramı şeklindeki karar ağaç yapısı şekil 4.4’de verilmiştir:

- Ara düğümler, bir nitelik sınamasını **Nit ?**
- Dallar, sınama sonucunu →
- Yapraklar, Sınıfları temsil eder. **Cl**



Şekil 4.4. Karar ağaç yapısı



Şekil 4.5. Tarım ürünleri karar ağaç yapısı[11],[12],[16],[17],[18]

Şekil 4.5’de her bir tarım ürünü için sınıflandırma aşamasında ortalama sıcaklık, min sıcaklık, max sıcaklık, nisbi nem, güneşlenme süresi, yağış toplamı, rüzgar hızı ve toplam sıcaklık olmak üzere 8 koşul baz alınmıştır.

Ürünün yetiştirilmesi için gerekli olan koşul sağlanıyorsa 0 parametresi, sağlanmıyorsa 1 parametresi ile temsil edilmiştir. Koşulu sağlamayan parametre sayısı 0 veya 1 ise verimli, 2 ise az verimli, 3 ise vegetatif, 3 den büyükse verimsiz olarak sınıflandırma yapılmıştır.

Tablo 4.12’de karar ağaç teknikleri ile ürünlerin verimli, az verimli, vegetatif ve verimsiz olarak önceden belirlenmiş sınıflardan hangisine dahil olduğu bulundu. Bu teknik ile küçük ağaçların yorumlanması kolaydır, ayrıca sürekli ve ayırık nitelik değerleri için kullanılabilir.

Tablo 4.12. Tarım ürünlerinin sınıflandırılması

SINIFLANDIRMA MODELİ								
TAHILLAR	BAKLAGİLLER	ENDÜSTRİYEL BİTKİLER	MEYVESİ YENENLER	YUMURSU YENİLENLER	YAPRAĞI YENİLENLER	YUMUŞAK ÇEKİRDEKLİLER	SERT ÇEKİRDEKLİLER	ÜZÜMSÜ BİTKİLER
Arpa	Bakla	Ay Çiçeği	Biber	Havuç	Dereotu	Anmut	Badem	Çilek
Buğday	Bamya	Fig	Domates	Patates	Enginar	Ayva	Ceviz	Dut
Çavdar	Barbunya	Haşhaş	Hıyar	Sarımsak	Ispanak	Elma	Erik	İncir
Çeltik	Bezelye	Kendir	Kabak	Soğan	Karnabahar	Limon	Fındık	Kivi
Dan	Börülce	Kenevir	Karpuz	Turp	Kereviz	Mandalina	Kayısı	Muz
Mısır	Fasulye	Keten	Kavun	Yer Fıstığı	Lahana	Muşmıla	Kestane	Nar
Yulaf	Mercimek	Pamuk	Patlıcan		Marul	Portakal	Kızılçak	Üzüm
	Nohut	Soya Fasulyesi			Maydonoz	Trabzon Hurması	Kiraz	
		Şeker Pancarı			Nane	Turunç	Şeftali	
		Tütün			Paz	Yeni Dünya	Vişne	
					Pırasa		Zerdali	
					Semiz Otu		Zeytin	
Verimli								
Az Verimli								
Vegetatif								
Verimsiz								

Uygulamada, Malatya da verimli olarak yetiştiği bilinen kayısı, kiraz ve buğday gibi ürünler ve hiç yetişmediği bilinen zeytin, muz ve limon gibi ürünler yardımıyla sınıflandırma modelinin doğruluğu sınanmıştır.

4.4. Modelin kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi başka bir uygulamanın alt parçası olabilir. Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri bulurlar.[8] Malatya da yapılan meteorolojik uygulamalardan elde edilen parametreler üzerinde kurulan ve uygulanan model Erzurum ve Iğdır illerinde yapılan meteorolojik uygulamalardan elde edilen parametreler üzerinde uygulandı(Tablo 4.14, Tablo 4.16). İlişkilendirme modeli ile Doğu Anadolu Bölgesinde yetişen tarımsal ürünlerin desen keşfinin yapılması sağlandı.

Tablo 4.13.Erzurum ili meteorolojik parametreler[11],[12],[14]

AYLAR	Ortalama Sıcaklık °C	Min Sıcaklık °C	Max Sıcaklık °C	Nisbi Nem %	Güneşlenme Süresi	Yağış Toplamı m ² /kg	Rüzgar Hızı Km/saat
OCAK	-9,6	-15,0	-4,1	77	2,8	20,0	2,1
ŞUBAT	-8,6	-14,1	-2,7	77	3,8	25,3	2,3
MART	-2,9	-8	2,5	75	4,8	31,3	2,6
NİSAN	5,4	-0,2	11,5	66	6	57,8	3,3
MAYIS	10,3	3,6	16,8	64	7,6	70,3	3,1
HAZİRAN	14,8	6,6	21,9	58	10	41	2,8
TEMMUZ	19,3	10,4	26,9	52	11,1	25,7	3,2
AĞUSTOS	19,3	10,1	27,6	50	10,8	14,6	2,9
EYLÜL	14,3	5,2	23,1	52	8,9	20,6	2,7
EKİM	7,5	0,7	15,3	65	6,5	48,3	2,5
KASIM	0,2	-5,1	6,5	73	4,3	31,8	2,3
ARALIK	-6,4	-11,3	-1,3	78	2,5	22,6	2,1

Tablo 4.14.Erzurum ili tarımsal ürün sınıflandırması

SINIFLANDIRMA								
TAHILLAR	BAKLAGİLLER	ENDÜSTRİYEL BİTKİLER	MEYVESİ YENENLER	YUMRUSU YENİLENLER	YAPRAÇI YENİLENLER	YUMUŞAK ÇEKİRDEKLİLER	SERT ÇEKİRDEKLİLER	ÜZÜMSÜ BİTKİLER
Arpa	Bakla	Ay Çiçeği	Biber	Havuç	Dereotu	Armud	Badem	Çilek
Buğday	Bamya	Fiğ	Domates	Patates	Enginar	Ayva	Ceviz	Dut
Çavdar	Barbunya	Hağuş	Hıyar	Sarımsak	Ispanak	Elma	Erik	İncir
Çeltik	Bezelye	Kendir	Kabak	Soğan	Karnabahar	Limon	Fındık	Kivi
Dan	Börülce	Kenevir	Karpuz	Turp	Kereviz	Mandalina	Kayısı	Muz
Mısır	Fasulye	Keten	Kavun	Yer Fıstığı	Lahana	Muşmula	Kestane	Nar
Yulaf	Mercimek	Pamuk	Patlıcan		Marul	Portakal	Kırmık	Üzüm
	Nokut	Soya Fasulyesi			Maydanoz	Trabzon Hurması	Kiraz	
		Şeker Pancarı			Nane	Turunc	Şeftali	
		Tütün			Pazı	Yeni Dünya	Vişne	
					Pırasa		Zerdali	
					Semiz Otu		Zeytin	
Verimli								
Az Verimli								
Vegetatif								
Verimsiz								

Tablo 4.15.Iğdır ili meteorolojik parametreler[11],[12],[14]

AYLAR	Ortalama Sıcaklık °C	Min Sıcaklık °C	Max Sıcaklık °C	Nisbi Nem %	Güneşlenme Süresi	Yağış Toplamı m ² /kg	Rüzgar Hızı Km/saat
OCAK	-2,9	-7,3	2,6	66	2,7	13,3	1,0
ŞUBAT	-9,0	-4,9	5,6	60	4,1	15,9	1,3
MART	6,3	0,1	12,8	51	5,4	21,5	1,6
NİSAN	13,4	6,5	20,1	50	6,1	37,1	1,6
MAYIS	17,4	10,5	24,3	51	7,5	48,9	1,4
HAZİRAN	22	14,3	29,4	46	9,6	33,8	1,5
TEMMUZ	26	18,3	33,7	44	10,1	13,1	1,5
AĞUSTOS	25,1	17,7	33,4	46	9,9	8,5	1,4
EYLÜL	19,8	12,2	29,1	61	8,5	9,8	1,2
EKİM	12,5	6,1	21,2	62	6,3	26,5	0,9
KASIM	5,4	0,2	12,6	65	4,3	18,4	0,9
ARALIK	-0,2	-4,2	4,9	68	2,4	12,3	0,9

Tablo 4.16.İğdır ili tarımsal ürün sınıflandırması

SINIFLANDIRMA								
TAHILLAR	BAKLA GİLLER	ENDÜSTRİYEL BİTKİLER	MEYVESİ YENENLER	YUMRUSU YENİLENLER	YAPRAĞI YENİLENLER	YUMUŞAK ÇEKİRDEKLİLER	SERT ÇEKİRDEKLİLER	ÜZÜMSÜ BİTKİLER
Arpa	Bakla	Ay Çiçeği	Biber	Havuç	Dereotu	Armut	Badem	Çilek
Buğday	Bamya	Fiğ	Domates	Patates	Enginar	Ayva	Ceviz	Dut
Çavdar	Barbunya	Haşhaş	Hıyar	Sarımsak	Ispanak	Elma	Erik	İncir
Çeltik	Bezelye	Kendir	Kabak	Soğan	Karnabahar	Limon	Fındık	Kivi
Dan	Böğülice	Kenevir	Karpuz	Turp	Kereviz	Mandalina	Kayısı	Muz
Mısır	Fasulye	Keten	Kavun	Yer Fıstığı	Lahana	Muşmıla	Kestane	Nar
Yulaf	Mercimek	Pamuk	Patlıcan		Marul	Portakal	Kızılçık	Üzüm
	Nohut	Soya Fasulyesi			Maydonoz	Trabzon Hurması	Kiraz	
		Şeker Pancarı			Naane	Turunç	Şeftali	
		Tütün			Pazı	Yeni Dünya	Vişne	
					Pırasa		Zerdali	
					Semiz Otu		Zeytin	
Verimli								
Az Verimli								
Vegetatif								
Verimsiz								

Apriori Algoritmasına İlişkin Varsayımlar :

Kullanılan market sepeti verisinde her harekette yer alan ürün kodları sayısaldır ve ürün kodları küçükten büyüğe doğru sıralıdır. Öge kümeleri eleman sayıları ile birlikte anılır ve k adet ürüne sahip bir öge k -öge kümesi ile gösterilir. Öge kümelerinde yer alan ürün kodları küçükten büyüğe sıralıdır. Her öge kümesine destek metriğini tutmak üzere bir sayaç değişkeni iliştilmiştir. Bu sayaç değişkeni öge kümesi ilk kez yaratıldığında sıfırlanır. Aday öge kümeleri C ile gösterilir ve k -öge kümesine ilişkin aday kümesi $c[1], c[2], c[3], \dots, c[k]$ ürünlerini içerir ve bu ürünler de $c[1] < c[2] < c[3] < \dots < c[k]$ olacak şekilde sıralıdır.

Algoritmada kullanılan değişkenler tablo 4.17’de özetlenmiştir. Apriori algoritması tablo 4.18’de verilmiştir.

Tablo 4.17.Apriori Algoritmasında Kullanılan Değişkenler

k-ögeküme	K adet öge içeren küme
Lk	Sık geçen k-öge küme kümesi (Bu kümeler minimum destek kistasını sağlarlar). Bu kümenin her üyesi iki alandan oluşur. i) öge kümesi ii) destek sayacı
Ck	Aday k-ögeküme kümesi (Bu kümeler potansiyel olarak sıkgeçen öge kümeleridir). Bu kümenin her üyesi iki alandan oluşur. i) öge kümesi ii) destek sayacı

Tablo 4.18’de verilen Apriori algoritmasında yer alan apriori-gen işlevi, $(k-1)$ adet ögeye sahip L_{k-1} ögeler kümesini kullanarak k adet ögeye sahip aday kümeleri üretir. Bu işlev şu biçimde çalışır. İlk önce, L_{k-1} ile L_{k-1} bitleştirme (*join*) işlemine tabi tutulur. Oluşan kümeler budanarak (*pruning*) işlevden dönülür. Budama işleminde c aday kümesinin $(k-1)$ ögeye sahip alt kümelerinden L_{k-1} ’de yer almayan kümeler silinir. Apriori-gen işlevinin algoritma kesiti, tablo 4.19’da SQL dilinden faydalanılarak verilmiştir.

Tablo 4.18. Apriori Algoritması Kesiti

```

L1 = { sık geçen 1-öge küme kümeleri }
SAYARAK YINELE ( k:=2, Lk-1 ≠ □, k:=k+1 ) [
/* k adet ögeye sahip aday kümelerin bulunması */
Ck = apriori-gen(Lk-1);
TUM t ∈ D hareketler için [
/* t hareketinde yer alan aday kümelerin bulunması */
Ct = subset(Ck, t);
TUM c ∈ Ct aday kümeler için
c.sayac := c.sayac+1;
]
Lk = { c ∈ Ck | c.sayac ≥ min-destek
}
Uk Lk DONDUR

```

Tablo 4.19. Apriori-gen Aday Küme Üretme Algoritma Kesiti

```

INSERT INTO Ck
SELECT p.öge1, p.öge2, ... p.ögek-1, q.ögek-1
FROM Lk-1 p, Lk-2 q
WHERE p.öge1 = q.öge1 and ... p.ögek-2 = q.ögek-2 and
p.ögek-1 <
q.ögek-1;
TUM c ∈ Ck aday kümeler için
TUM c kümesinin (k-1) ögeye sahip tüm alt kümeleri için
EĞER ( s ∈ Lk-1 ) İSE
DELETE c FROM Ck;

```

İlişkilendirme kuralının oluşturulması;

Tablo 4.20.Tarım ürünleri verisi

İller	Yetişen Tarım Ürünleri
Malatya	Arpa, Buğday, Çavdar, Fasulye, Nohut, Mercimek, Pancar, Tütün, Biber, Domates, Hıyar, Maydanoz, Armut, Elma, Erik, Kayısı, Kiraz, Vişne, Zerdali, Çilek, Dut
Erzurum	Buğday, Çavdar, Biber, Domates, Hıyar, Pancar, Elma, Erik, Kiraz, Vişne
Iğdır	Arpa, Buğday, Çavdar, Pancar, Tütün, Biber, Domates, Hıyar, Armut, Elma, Erik, Kayısı, Kiraz, Vişne, Şeftali

İlişki tanımlamada güven (confidence) ve destek (Support) olarak iki değer kullanılır.

Support($X \rightarrow Y$) : X ve Y item setini içeren hareketlerin sayısı

$$\text{Support}(A \rightarrow B) = \text{Support}(A \cup B)$$

Confidence($X \rightarrow Y$) : X iteminin bulunduğu yerde X ve Y item setinin bulunma olasılığı

$$\text{Confidence}(A \rightarrow B) = \text{Support}(A \cup B) / \text{Support}(A)$$

Adım 1. Minimum destek sayısı (min.support) ve minimum güven değeri (min.confidence) belirlenmesi.

$$\text{min.support} = 1$$

$$\text{min.confidence} = \%80$$

Adım 2. Item setler içerisindeki her bir itemin destek değerinin bulunması.

Tablo 4.21. İtemlerin destek değerleri

İtem	Destek değ.
Buğday	3
Arpa	2
Çavdar	3
Fasulye	1
Nohut	1
Mercimek	1
Pancar	3
Tütün	2
Biber	3
Domates	3
Hıyar	3
Maydanoz	1
Armut	2
Elma	3
Erik	3
Kayısı	2
Kiraz	3
Vişne	3
Zerdali	1
Çilek	1
Dut	1
Şeftali	1

Adım 3. Minimum destek değerinden daha küçük desteğe sahip olan itemlerin devre dışı bırakılması.

Tablo 4.22. Tekli birlikteliklerin min.destek değerine göre düzenlenmesi

Item	Destek değ.
Buğday	3
Arpa	2
Çavdar	3
Pancar	3
Tütün	2
Biber	3
Domates	3
Hıyar	3
Armut	2
Elma	3
Erik	3
Kayısı	2
Kiraz	3
Vişne	3

Adım 4. Elde edilen tekli birliktelikler dikkate alınarak ikili birlikteliklerin oluşturulması.

Tablo 4.23. İkili birlikteliklerin oluşturulması

Item	Destek değ.
Buğday - Çavdar	3
Buğday - Arpa	2
Buğday - Pancar	3
Buğday - Tütün	2
Buğday - Biber	3
Buğday - Domates	3
Buğday - Hıyar	3
Buğday - Armut	2
Buğday - Elma	3
Buğday - Erik	3
Buğday - Kayısı	2
Buğday - Kiraz	3
Buğday - Vişne	3

Adım 5. İkili birlikteliklerin oluşturulması

Min.destek değeri 2 olarak seçilirse ve bu değerden daha düşük destek değerine sahip olan item setler dikkate alınmazsa liste tablo 4.24 deki gibi olur.

Tablo 4.24. İkili birlikteliklerin min.destek değerleri dikkate alınarak düzenlenmesi

Item	Destek değ.
Buğday - Çavdar	3
Buğday - Pancar	3
Buğday - Biber	3
Buğday - Domates	3
Buğday - Hıyar	3
Buğday - Elma	3
Buğday - Erik	3
Buğday - Kiraz	3
Buğday - Vişne	3

Adım 6. Üçlü birlikteliklerin oluşturulması

Tablo 4.25. Üçlü birlikteliklerin oluşturulması

Item	Destek değ.
Buğday - Çavdar - Pancar	3
Buğday - Çavdar - Biber	3
Buğday - Çavdar - Domates	3
Buğday - Çavdar - Hıyar	3
Buğday - Çavdar - Elma	3
Buğday - Çavdar - Erik	3
Buğday - Çavdar - Kiraz	3
Buğday - Çavdar - Vişne	3

Adım 7. Üçlü birlikteliklerin her birinin birliktelikler kuralının oluşturulması

Buğday, Çavdar, Pancar için birliktelikler kuralının alt kümesi ;
{Buğday, Çavdar}, {Buğday, Pancar}, {Çavdar, Pancar}, {Buğday}, {Çavdar}, {Pancar}

Tablo 4.26. Buğday, Çavdar ve Pancar için birliktelik kurallarının alt kümesi

Birliktelik	Açıklama	Güven
Buğday & Çavdar -> Pancar	Buğday ve Çavdarın bulunduğu item sette Pancarın olma olasılığı	$3/3 = \%100$
Buğday & Pancar-> Çavdar	Buğday ve Pancarın bulunduğu item sette Çavdarın olma olasılığı	$3/3 = \%100$
Pancar & Çavdar -> Buğday	Çavdar ve Pancarın bulunduğu item sette Buğdayın olma olasılığı	$3/3 = \%100$
Buğday -> Pancar & Çavdar	Buğdayın bulunduğu item sette Çavdar ve Pancarın bulunma olasılığı	$3/3 = \%100$
Çavdar -> Buğday & Pancar	Çavdarın bulunduğu item sette Buğday ve Pancarın bulunma olasılığı	$3/3 = \%100$
Pancar -> Buğday & Çavdar	Pancarın bulunduğu item sette Buğday ve Pancarın bulunma olasılığı	$3/3 = \%100$

Bu birliktelikte min.confidence (% 80) değeri dikkate alındığında ;
Buğday ve Çavdarın yetiştiği yerde Pancar yetiştirme olasılığı % 100 dür.
Buğday ve Pancarın yetiştiği yerde Çavdar yetiştirme olasılığı % 100 dür.
Pancar ve Çavdarın yetiştiği yerde Buğday yetiştirme olasılığı % 100 dür.
Buğdayın yetiştiği yerde Çavdar ve Pancar yetiştirme olasılığı % 100 dür.
Çavdarın yetiştiği yerde Buğday ve Pancar yetiştirme olasılığı % 100 dür.
Pancarın yetiştiği yerde Çavdar ve Buğday yetiştirme olasılığı % 100 dür.

Tüm ürünlerin Apriori algoritması ile analiz edilmesi sonucu;

Verimli: Buğday, Çavdar, Pancar, Biber, Domates, Hıyar, Elma, Erik, Kiraz, Vişne

Az Verimli: Fig, Soğan

Vegatatif: Bamya, Havuç, Turp, Fındık

Verimsiz: Çeltik, Darı, Yer Fıstığı, Kereviz, Limon, Mandalina, Portakal, Turunç, Yeni Dünya, Kivi, Muz ilişkisi ortaya çıkmıştır.

Tablo 4.27.Doğu Anadolu Bölgesi tarımsal ürün desen keşfi

İLİŞKİLENDİRME								
TAHILLAR	BAKLA GİLLER	ENDÜSTRİYEL BİTKİLER	MEYVESİ YENENLER	YUMRUSU YENİLENLER	YAPRAĞI YENİLENLER	YUMUŞAK ÇEKİRDEKLİLER	SERT ÇEKİRDEKLİLER	ÜZÜMSÜ BİTKİLER
Arpa	Bakla	Ay Çiçeği	Biber	Havuç	Dereotu	Armut	Badem	Çilek
Buğday	Bamya	Fiğ	Domates	Patates	Enginar	Ayva	Ceviz	Dut
Çavdar	Barbunya	Haşhaş	Hıyar	Sarımsak	Ispanak	Elma	Erik	İncir
Çeltik	Bezelye	Kendir	Kabak	Soğan	Karnabahar	Limon	Fındık	Kivi
Dan	Börülce	Kenevir	Karpuz	Turp	Kereviz	Mandalina	Kayısı	Muz
Mısır	Fasulye	Keten	Kavun	Yer Fıstığı	Lahana	Muşmula	Kestane	Nar
Yulaf	Mercimek	Pamuk	Patlıcan		Marul	Portakal	Kızılçık	Üzüm
	Nohut	Soya Fasulyesi			Maydonoz	Trabzon Hurması	Kiraz	
		Şeker Pancarı			Nane	Turunç	Şeftali	
		Tütün			Pazı	Yeni Dünya	Vişne	
					Pırasa		Zerdali	
					Semiz Otu		Zeytin	
Verimli								
Az Verimli								
Vegetatif								
Verimsiz								

4.5 Modelin İzlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesi ve gerekiyorsa yeniden düzenlenmesi gerekecektir. Uygulamada gerek yeni tarım ürünlerinin sisteme dahil edilmesi gerekse daha sonraki yıllarda yapılacak meteorolojik parametrelerin sisteme dahil edilmesi sağlandı. Ayrıca tahmin edilen ve gözlenen değişiklikler farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanıldı.

5.SONUÇ

5.1.Sonuçlar

Bu tez çalışmasında veri madenciliği modellerinden kümeleme, sınıflandırma ve ilişkilendirme modelleriyle işlenmiş bunun sonucunda Malatya'nın ziraî meteoroloji haritası çıkartılmış ve Doğu Anadolu Bölgesi tarımsal ürün desen keşfinin yapılması sağlanmıştır.

Tezde ilk olarak veri madenciliği açıklanmış ve veri madenciliğinde kullanılan modeller hakkında kısaca bilgi verilmiştir. Gerçekleştirilen yazılımda ise veri madenciliği yöntemleri kullanılarak Malatya'nın ziraî meteoroloji haritası çıkartılmıştır. Ayrıca yazılımın seçilen pilot iller üzerinde uygulanmasıyla Doğu Anadolu Bölgesi tarımsal ürün desen keşfi yapılmıştır.

Büyük boyutta verinin saklı olduğu meteoroloji ve tarımsal veri tabanlarının bireysel çalışmalar ile işlenmesi ve yorumlanması neredeyse imkansızdır. Tezde Microsoft SQL Server 2005 Home ve Visual Basic 6.0 programları kullanılarak büyük veri yığınlarının işlenip yorumlanması mümkün hale gelmiş klasik yöntemlerle bulunması çok zor veya imkansız olan ilişkilerde ortaya çıkarılmıştır.

Tezde ortaya çıkan bazı ilginç ilişkiler; Malatya da tarım alışkanlığı olarak kazanılmamış olan yulaf, fiğ, haşhaş, zerdali ve nar'ın yetişebileceği görülmüştür. Ayrıca hiç yetişmediği düşünülen şeftali, kıvılcık, fındık, enginar, havuç, turp, pamuk, ayçiçeği, bakla, bamya, börülce ve mısır ürünlerinin vegatatif olarak yetişebileceği saptanmıştır.

Bulunan bu bilgiler ile üreticinin yapacağı yatırımı önceden planlaması böylelikle yatırımdan maksimum faydayı sağlaması yanlış planlamadan dolayı oluşacak zararın minimuma indirilmesi bu bağlamda yapılacak büyük projelerde ülke ekonomisine katkıda bulunulması sağlanmıştır.

Bir bütün olarak değerlendirme yapıldığında yazılımın basit, kullanılabilir ve sınama verisi veya yeni veriler üzerinde geçerli olduğu görülmüştür.

5.2. Öneriler

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde değişiklikler ortaya çıkabilir. Bilimim hızlı ilerlemesi göz önüne alındığında yazılımın çok uzun olmayan aralıklarla güncellemeleri yapılmalıdır.

Yazılım tarım ürünlerinin iklimsel istek değerleri göz önüne alınarak gerçekleştirilmiştir. Yükselti ve toprak gibi faktörler baz alınarak yazılım genel anlamda genişletilebilir.

KAYNAKLAR

- [1] www.bilgiyonetimi.org Eker,H;Veri Madenciliği veya Bilgi Keşfi,Ekim,2004
- [2] E.Kaya ve ark.,Tıpta Veri Ambarları Oluşturma ve Veri Madenciliği Uygulamaları, (Selçuk Ü.) , Akademik Bilişim 2003, Çukurova Üniversitesi, Adana, 3 şubat 2003
- [3] Zhi-Hua Zhou, Three Perspectives of Data Mining, Artificial Intelligence, 143(2003), p: 139-146
- [4] Methodologies for Knowledge Discovery and Data Mining : Third Pacific-Asia Conference, Pakdd-99, Beijing, China, April 26-28, 1999 : Proceedings, Zhong, N. - Zhou, L., Springer Verlag, 1999.
- [5] Data Mining Concepts and Techniques, Han, J.-Kamber,Kaufmann Publishers, 1st Ed., San Francisco, USA, 2000.
- [6], Han, J., An Introduction to Cluster Analysis for Data Mining, 2000 http://www-users.cs.umn.edu/~han/dmclass/cluster_survey_10_02_00.pdf
- [7] Claude Seidman, Data Mining with Microsoft SQL Server 200 Technical Refence. IT Professional and Developer R. Agrawal, T. Imielinski, A. Swami , 1993 , Mining Associations between Sets of Items in Massive Databases
- [8] Mining Multiple-Level Association Rules in Large Databases, Han J. - Fu Y.,IEEE Transactions on Knowledge and Data Engineering, vol 11 no.5, 1999
- [9] Veri Tabanlarında Bilgi Kesfi ve Veri Madenciligi, Akpınar, H., _stanbul Üniv._sletme Fakültesi Dergisi, C:29 S: 1 Nisan 2000
- [10] PC Chair, [Workshop on Data Mining and Business Intelligence](http://www3.itu.edu.tr/~sgunduz/courses/verimaden/index.html), (DMBI 2007), Istanbul, Turkey, April 2007.<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/index.html>
- [11] Yalçın. S, Zirai Meteoroloji, Devlet Meteoroloji işleri genel müdürlüğü, 2001
- [12] Türkiye İklim Bölgeleri ve Haritası, Mızrak.G, Ankara: Ota Anadolu Bölge Zîrai araştırma enstitüsü tarla bitkileri ıslahı bölümü, 1983, 24s. Ota Anadolu Bölge Zîrai araştırma enstitüsü genel yayın No:52, Teknik yayınları No:2.
- [13] Malatya Meteoroloji Bölge Müdürlüğü, Klimatolojik Rasat Cetveli, Malatya, 1971-2006
- [14]Devlet Meteoroloji İşleri Genel Müdürlüğü Bülteni, Ankara, 1971-2006
- [15] Fu Kanwalpreet, Yongjian.,Shih, Sandhu Ming-Yi, Clustering of Web Users Based on Access Patterns”, In Proceedings of WEBKDD, 1999.
- [16] Soylu, A.; Türk, R.; Genel Meyvecilik; Anadolu Ün. Yay.Eskişehir. 2002
- [17] Şeniz, V.,Özgür, M., Sivritepe, Ö., Özer. H; Sebzeçilik; Anadolu Ün. Yay.Eskişehir 2002
- [18] Yürür, N., Açıkgöz, E., Özgümüş, A., Azkan, N., Turan, M., Çelik, N., Yağdı, K., Ayçiçeği, M., Doğan, R., Turgut, İ., Göksoy, A., Uzun, A., Dağüstü, N., Kaçar, O.;Tarla Bitkileri; Anadolu Ün. Yay.Eskişehir 2002

ÖZGEÇMİŞ

1980-Elazığ doğumlu olan Fatih TOPALOĞLU ilk, orta ve lise öğrenimini Malatya’da tamamladı. 1998 yılında girdiği Fırat Üniversitesi Teknik Eğitim Fakültesi Bilgisayar ve Elektronik Eğitimi Bölümünden 2002 yılında mezun oldu. Aynı yıl Malatya Lisesinde göreve başladı. 2003 yılında kısa dönem er olarak askerlik görevini yerine getirdi. 2005 yılında Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim dalında yüksek lisans yapma hakkı kazanmıştır. Halen Malatya Ticaret Meslek Lisesinde bilgisayar öğretmeni olarak görev yapmaktadır.