



**BELEDİYELEDE SOSYAL YARDIM HİZMETLERİ DAĞITIMINA
YÖNELİK KARAR DESTEK SİSTEMİ GELİŞTİRİLMESİ**

Büşra TAN

**Yüksek Lisans Tezi
Yazılım Mühendisliği Anabilim Dalı
Danışman: Doç. Dr. Murat KARABATAK
OCAK-2018**

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BELEDİYELERDE SOSYAL YARDIM HİZMETLERİ DAĞITIMINA YÖNELİK KARAR
DESTEK SİSTEMİ GELİŞTİRİLMESİ

YÜKSEK LİSANS TEZİ

Büşra TAN

142137104

Tezin Enstitüye Verildiği Tarih: 2 Ocak 2018
Tezin Savunulduğu Tarih: 19 Ocak 2018

Tez Danışmanı: Doç. Dr. Murat KARABATAK (F.Ü)
Diğer Jüri Üyeleri: Prof. Dr. Engin AVCI (F.Ü)
Yrd. Doç. Dr. Ahmet Arif AYDIN (İ.Ü)

OCAK-2018

ÖNSÖZ

“Belediyelerde Sosyal Yardım Hizmetleri Dağıtımına Yönelik Karar Destek Sistemi Geliştirilmesi” adlı çalışmamın ortaya çıkmasında emeği geçen, çalışmamı büyük bir titizlikle inceleyen ve hoşgörüsünü göstermekten kaçınmayan Kıymetli Hocam Doç. Dr. Murat KARABATAK’ a teşekkür ederim. Ayrıca tez çalışma süresince akademik bilgi ve tecrübelerinden istifade ettiğim değerli dayım Yrd. Doç. Dr. Hulusi BİNBAŞIOĞLU’na, deneyimlerinden faydalandığım ve manevi desteğini esirgemeyen arkadaşım Arş. Gör. Elif VAROL’a, yüksek lisans yaptığım süreçte beni teşvik eden ve destekleyen Yeşilyurt Belediye Başkanı Sayın Hacı Uğur POLAT’a, tez çalışmama farklı şekillerde katkıda bulunan çalışma arkadaşlarım İsmihan ATEŞ ve Semra ÖZCAN’a ve hayatımın her anında desteği ile yanımda olan aileme teşekkür ederim.

Büşra TAN
ELAZIĞ, 2018

İÇİNDEKİLER

ÖNSÖZ	II
İÇİNDEKİLER	III
ÖZET	V
SUMMARY	VI
ŞEKİLLER LİSTESİ	VII
TABLolar LİSTESİ	VIII
KISALTMALAR LİSTESİ	IX
1. GİRİŞ	1
1.1 Literatür	2
1.2 Tezin Amacı	4
2. VERİ MADENCİLİĞİ	5
2.1 Veriyi Bilgiye Dönüştürme	5
2.2 Uygulama Alanları	6
2.2.1 Pazarlama	6
2.2.2 Bankacılık	6
2.2.3 Sigorta	6
2.2.4 Elektronik Ticaret	6
2.3 Veri Madenciliği Süreci	6
2.3.1 Veri Temizleme	7
2.3.2 Veri Bütünleştirme	7
2.3.3 Veri İndirgeme	8
2.3.4 Veri Dönüştürme	8
2.3.5 Veri Madenciliği Algoritmasının Uygulanması	9
2.3.6 Sonuçları Sunum ve Değerlendirme	9
2.4 Veri Madenciliği Yöntemleri	9
2.4.1 Sınıflandırma	9
2.4.2 Kümeleme	10
2.4.3 Birliktelik Kuralı	10
3. SINIFLANDIRMA	12
3.1 Model Başarım Ölçütleri	12
3.2 Sınıflandırma Algoritmaları	13
3.2.1 Karar Ağacı ile Sınıflandırma	14
3.2.2 ID3 Algoritması	15
3.2.3 C4.5 Algoritması	16
3.2.4 J48 Algoritması	17
3.2.5 K-En Yakın Komşu Algoritması	17
3.2.6 Bayes Sınıflandırıcısı	18
3.2.7 Yerel Ağırlıklı Öğrenme Algoritması	21
3.2.8 Sıralı Minimal Optimizasyon Algoritması	22
3.2.9 Yerine Koyarak Örnekleme Algoritması	23
3.2.10 Rastgele Orman Sınıflandırma Algoritması	24
4. SOSYAL YARDIM VERİ KÜMESİ İLE SINIFLANDIRMA UYGULAMASI	25
4.1 Veri Kümesi	25
4.2 Veri Kümesi Üzerinde Sınıflandırma Uygulamaları	27

4.3	Veri Kümesi Sınıflandırma Sonuçları.....	28
5.	GELİŞTİRİLEN KARAR DESTEK SİSTEMİ.....	30
5.1	Karar Destek Sisteminin Amacı.....	30
5.2	Karar Destek Sisteminin Planlanması.....	30
5.3	Karar Destek Sisteminin Çözümlemesi	33
5.4	Karar Destek Sistemini Tasarlama ve Geliştirme	35
5.5	Karar Destek Sisteminin Test ve Bakımı	39
6.	SONUÇLAR VE DEĞERLENDİRME	41
7.	KAYNAKLAR	42



ÖZET

Ülkemizde nüfusun artmasıyla meydana gelen işsizlik, yoksulluk vb. sorunlar belediyelerin sosyal yardım ve sosyal hizmet alanında görev üstlenmesine neden olmuştur. Bu amaçla sosyal yardım ve sosyal hizmet görevini üstlenen belediyelerin, sosyal yardım dağıtımını esnasında eldeki verileri hızlı bir şekilde kullanabilmesi ve ilgili komisyonlarda doğru kararlar verebilmesi büyük önem arz etmektedir.

Bu tez çalışmasında, Malatya Yeşilyurt Belediyesi örneği üzerinden sosyal yardım hizmetleri dağıtımına yönelik bir karar destek sisteminin geliştirilmesi amaçlanmıştır. Bu doğrultuda öncelikle Yeşilyurt Belediyesinden veriler elde edilmiş ve bu verilere veri madenciliği yöntemlerinden olan sınıflandırma algoritmaları uygulanarak elde edilen başarımlar karşılaştırılmıştır. Karşılaştırma sonucunda, ID3, C4.5, J48, K-En Yakın Komşu, Sade Bayes, Yerine Koyarak Örnekleme, Rastgele Orman, Sıralı Minimal Optimizasyon ve Yerel Ağırlıklı Öğrenme sınıflandırma algoritmaları en iyi sonuçları veren algoritmalar olarak tespit edilmiştir. En iyi sonucu veren üç farklı sınıflandırma algoritmasına ait sonuçlar ve kurallar kullanılarak, Yeşilyurt Belediyesi örneği üzerinde web tabanlı bir karar destek sistemi tez kapsamında gerçekleştirilmiştir.

Anahtar Kelimeler: Sosyal Yardım, Sosyal Hizmet, Veri Madenciliği, Sınıflandırma, Karar Destek Sistemi.

SUMMARY

Development of Decision Support System for Distribution of Social Assistance Services in Municipalities

Unemployment, poverty and similar problems that have come to fore with the increase in population in our country have caused the municipalities to take charge in the field of social assistance and social services. For this purpose, it is very important that the municipalities that undertake social assistance and social service tasks are able to use the present data quickly during the distribution of social assistance and make the right decisions in the related commissions.

In this thesis study, it is aimed to develop a decision support system for distributing social welfare services through the example of Malatya Yeşilyurt Municipality. In this direction, firstly data was obtained from Yeşilyurt Municipality and the results obtained were compared by applying classification algorithms which are data mining methods. As a result of comparison, ID3, C4.5, J48, K-Nearest Neighbor, Naive Bayes Classifier, Bootstrap Aggregating, Random Forest Algorithms, Sequential Minimal Optimization and Local Weighted Learning algorithms have been determined as algorithms that give the best results. A web-based decision support system on the Yeşilyurt Municipality case was carried out within the scope of the thesis, using the results and rules of three different classification algorithms giving the best result.

Keywords: Social Assistance, Social Services, Data Mining, Classification, Decision Support System

ŞEKİLLER LİSTESİ

Sayfa No

Şekil 2.1. Veri Madenciliği Süreci	7
Şekil 3.1. X ve Y nitelikleri üzerine uygulanan teste dayanan bir basit karar ağacı ..	14
Şekil 3.2. Gözlemlenen ▲ noktasına en yakın k=3 komşunun belirlenmesi.....	17
Şekil 3.3. T ₁ , T ₂ ve A olayları.....	19
Şekil 4.1. Arff Formatı Gösterimi.....	27
Şekil 4.2. Yerel Ağırlıklı Öğrenme Algoritmasının WEKA aracında Gösterimi.....	29
Şekil 5.1. Karar Destek Sistemi Zaman-İş Planı Tablosu.....	31
Şekil 5.2. Karar Destek Sisteminde Admin Kullanıcı İşlemleri.....	33
Şekil 5.3. Karar Destek Sisteminde Normal Kullanıcı İşlemleri.....	34
Şekil 5.4. Karar Destek Sistemi Kullanım Durum Diyagramı	34
Şekil 5.5. Karar Destek Sistemi Veri Modeli	35
Şekil 5.6. Panel Giriş Ekranı.....	35
Şekil 5.7. Veri Giriş Ekranı	36
Şekil 5.8. Aile Bilgileri Giriş Ekranı	37
Şekil 5.9. Özet Form Ekranı	38
Şekil 5.10. Yardım Sonuç Ekranı.....	39

TABLolar LİSTESİ

	<u>Sayfa No</u>
Tablo 2.1. Beş (5) Müşterinin sepetindeki ürünler.....	10
Tablo 3.1. İki sınıf için karışıklık matrisi.....	12
Tablo 4.1. Anket soruları ve cevapları.....	26
Tablo 4.2. Sınıflandırma sonuçları.....	28
Tablo 4.3. YAÖ Algoritması karışıklık matrisi.....	28
Tablo 4.4. YAÖ Algoritması doğruluk çizelgesi	29
Tablo 5.1. Karar Destek Sistemi Genel Sistem Puanlama Tablosu	31

KISALTMALAR LİSTESİ

VTYS	: Veri Tabanı Yönetim Sistemi
VTBK	: Veri Tabanında Bilgi Keşfi
SSDM	: Sosyal Güvenlik Veri Madenciliği
VBK	: Veri Tabanında Bilgi Keşfi
DNO	: Doğru Negatif Oranı
YNO	: Yanlış Negatif Oranı
DPO	: Duyarlılık veya Doğru Pozitif Oranı
YPO	: Yanlış Pozitif Oranı
KNN	: K-En Yakın Komşu
BS	: Sade Bayes Sınıflandırıcı
YKÖ	: Yerine Koyarak Örnekleme
ROS	: Rastgele Orman Algoritmaları
SMO	: Sıralı Minimal Optimizasyon
YAO	: Yerel Ağırlıklı Öğrenme
SVM	: Destek Vektör Makineleri
QP	: Kuadratik Programlama

1. GİRİŞ

Çağın değişimiyle birlikte, günlük yaşam aktivitelerinin her parçasında bilgi dağıtımı gerçekleşmesi günümüzde “Bilgi Çağı” olarak adlandırılmaktadır [1]. Bilgi çağının hızla gelişimiyle, teknolojik araçlar yaşamın her alanında kullanılmaya başlanmış ve bu kullanımlar sonucunda oluşan bilgiler çok büyük veri yığınlarına sebep olmuştur. Bu verilere yani veri yığınlarına internetin olduğu her yerden ulaşılabilir. Fakat günümüzde bilgiye her alandan ulaşmak yerine ulaşılan bilgilerden maksimum sonuç çıkarmak daha önemli hale gelmiştir. Dolayısıyla bu tür bir durum, veri tabanı sistemleri ve Veri Madenciliği kavramlarını daha önemli bir hale getirmiştir.

Veri tabanı, seçilen veri ya da kayıtları sistemli olarak depolayabilen, güncelleyebilen, yönetebilen, taşıyabilen ve anlamlandırabilen düzenli bilgiler topluluğu olarak adlandırılmaktadır [2,3]. Tüm bilgilerin kaydedilmesine ve erişilmesine olanak sağlayan yazılım teknolojilerine yani veri tabanının yönetim veya kontrolünün oluşumuna Veri Tabanı Yönetim Sistemi (VTYS) denir. Veri Tabanı Yönetim Sisteminde bir veri girişi ve depolama, klasik dosyalamalardan farklıdır. Klasik dosyalama kullanılırken dosya yapısı veya kaydı için gerçekleştirilen herhangi bir değişiklik uygulama programının tamamının değişimine ve tekrar derlemeye sebep olmaktadır [3]. Dolayısıyla, geleneksel sorgu ve raporlama araçları, oldukça büyük olan veri yığınları arasında yetersiz kalmaktadır. Bu nedenle Veri Tabanlarında Bilgi Keşfi (VTBK) adı altında sürekli ve yeni arayışlar ortaya çıkmaktadır. VTBK süreci içerisinde modelin kurulması ve değerlendirilmesi aşamalarında meydana gelen Veri madenciliği, en önemli kısmı oluşturmaktadır ve bu sebepten, birçok araştırmacı tarafından VTBK ve veri madenciliği terimleri eş anlamlı olarak kullanılmaktadır [1].

Örneğin bir markette, her ürün için bir sonraki ayın satış öngörülerini çıkarılarak veri analizi yapılabilir ve müşterilerin satın aldıkları ürünlere bağlı olarak gruplanabilir. Böylece yeni bir ürün için potansiyel müşteriler belirlenerek, müşterilerin zaman içindeki hareketleri incelenebilir ve çeşitli öngörüler yapılabilir. Binlerce ürünün ve müşterinin olabileceği düşünüldüğünde bu analizin gözle veya elle yapılamayacağı, otomatik olarak yapılmasının gerektiği ortaya çıkmaktadır. İşte bu aşamada veri madenciliği devreye girmektedir.

Kısaca veri madenciliği; büyük ölçekli veriler arasından ‘değeri olan’ bir bilgiyi elde etme işidir. Bu sayede veriler arasındaki ilişkileri ortaya koymak ve gerektiğinde de ileriye yönelik kestirimlerde bulunmak mümkün görülmektedir.

Veri madenciliği; makine öğrenmesi, yapay zekâ, istatistik, örüntü tanıma gibi yöntemler kullanılarak anlamlı veriler ve kurallar çıkarmaktır. Yakın geleceğin de geçmişten çok fazla farklı olmayacağını varsayarsak, çıkarılmış olan bu kurallar gelecekte de geçerli olacak ve ilerisi için doğru öngörüler yapılmasını sağlayacaktır.

1.1 Literatür

Ülkemizde kapsamlı bir uygulama alanına sahip olan veri madenciliği pazarlama, bankacılık, sigortacılık, tıp ve e-ticaret gibi birçok alanda kullanılmaktadır [4]. Veri madenciliği pazarlama alanı olarak, müşterinin satın alırken davranış biçimlerinin belirlenmesi, alışveriş sepeti analizinin yapılması ve müşteri değerlendirilmesi vb. kullanılmaktadır. Bankacılık alanında, kredi kartı dolandırıcılığının ve sahteciliğinin belirlenmesi, kredi kartı ödemelerine dayalı müşteri gruplarının belirlenmesi vb. kullanılmaktadır. Sigortacılık alanında ise potansiyel müşterilerin tahmini vb. kullanım alanına sahiptir.

Karabatak ve İnce, birliktelik kuralı algoritmalarından apriori algoritması kullanarak öğrenciler üzerinde bir çalışma yapmış ve öğrencilerin başarılarına yönelik bir analiz gerçekleştirmiştir [5]. Koyuncuoglu ve Özgülbaş, veri madenciliğinin tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları üzerine bir çalışma yaparak sağlık sektöründe veri madenciliği kullanımına farklı bir bakış açısı kazandırmıştır [6]. Emel ve Taşkın, veri madenciliğinde karar ağacı algoritmalarından olan C&RT algoritmasından faydalanılarak perakendeci bir işletmenin müşterilerine göre satış davranışlarını içeren bir satış analizi uygulaması yapmıştır [7]. Mocan ve Büyüklü, veri madenciliğinin perakende sektöründe kullanımı üzerine çalışma yaparak kapsamlı veri setinden anlamlı bilgi çıkarımını sağlayacak perakende firmalarına yönelik bir araştırma yapmıştır [8]. Sadıqmal ve Yurtay, veri madenciliğinde birliktelik kuralı algoritmalarından olan apriori algoritmasını birtakım tıpla ilgili verilere uygulayıp, modern tıp bilgilerinin saklandığı veri tabanından faydalı bilgiler oluşturarak hekimlere yardımcı olacak bir karar destek sistemi geliştirilmesini göstermiştir [9].

Birçok alanda olduğu gibi sosyal yardım hizmetleri alanında da çok sayıda sayısal veri depolanmakta ve bu verilerin bilgiye dönüştürülmesinde veri madenciliği teknikleri

kullanılabilmektedir. Türkiye’de sosyal yardım gereksiniminin sağlanması anayasal hak olarak oluşturulduğundan beri devletin sosyal bir devlet olduğu anayasanın ikinci maddesinde belirtilmesiyle Türkiye’de yapılan sosyal yardım uygulamaları farklı şekillerde gerçekleşmektedir [10]. Muhtaç, kimsesi olmayan 65 yaşını dolduran vatandaşlara aylık bağlanması kanunu 1 Ocak 1977 tarihinde uygulanmaya başlamış olup aynı zamanda 18 yaşından büyük özürllülere ve 18 yaşını bitirmemiş özürllü yakını olan Türk vatandaşlarına aylık bağlanmaktadır. İlköğretim ve ortaöğretimde okuyan yardıma muhtaç öğrencilere burslu ve ücretsiz yatılı okuma olanağı sağlama görevi 17 Haziran 1982 tarihinde Milli Eğitim Bakanlığı’nca sağlanmıştır. 14 Haziran 1986 tarihinde Başbakanlığa bağlı Sosyal Yardımlaşma ve Dayanışma Genel Müdürlüğü Kanunu uygulanmıştır [10].

2005 yılında kanunlaşan 5393 Sayılı Belediye Yasası ve 5216 Sayılı Büyükşehir Belediyesi Yasası’yla muhtaç ve gereksinim sahiplerine sağlanacak sosyal yardımlar oluşturulmuştur. Belediyeler kanunun değişimiyle beraber belediyelere daha öncekilerden farklı bir misyon ve sorumluluk verilmesi bu aşamayı hızlandırmıştır. Kadın ve çocukları koruma haneleri açma vazifesi Büyükşehir Belediyelerine ve nüfusu 50.000’i geçen belediyelere verilmiştir. Yasanın 60. Maddesinde düşük kazançlı, kazancı bulunmayan ile özürllülere sağlanacak sosyal hizmet ve yardımlar belediye bütçesinden karşılanacak giderlerden sayılmıştır. Yasanın 38.maddesinde ise, Belediye başkanının vazifeleri arasında gelir ve giderlerin tümünden fakir ve kazancı bulunmayanlar için ayrılan ödeneği kullanmak, özürllülere hizmet sağlamak ve özürllü merkezleri açmak olarak belirtilmiştir [10, 11].

Cao, sosyal güvenlik ve sosyal refah verilerinin incelenmesine ilişkin önemli bir çalışma rapor edilmemesinden dolayı bu verilerle ilgili çok sayıda referansın kapsamlı bir araştırmasını yapmış ve büyük miktarda sosyal güvenlik ve sosyal refah verileri toplamıştır. Bu sayede iş ve araştırma konuları, sosyal güvenlik / refah hizmetleri ve verileri, sosyal güvenlik / refah verilerinin incelenmesindeki zorluklar, hedefler ve görevler de dâhil olmak üzere bir Sosyal güvenlik veri madenciliği (SSDM) çerçevesi oluşturmuştur [13].

Kum, Duncan, Flair ve Wank, Sosyal refah devlet kurumlarının idari veri tabanlarından toplumda en çok ihtiyacı olan ailelerin ve bireylerin deneyimleri hakkındaki bilgileri alarak bu bilgileri veri madenciliğinin doğru bir şekilde uygulanmasıyla anlamlı bilgi çıkarıp analiz etmiştir [14].

Anderson ve Mansingh, veri madenciliği kullanarak sosyal refah alanındaki verilerden desen çıkarabilmeyi araştırmıştır [15].

1.2 Tezin Amacı

Belediyeler halkın ihtiyaçlarını karşılamak adına, yerel ve bölgesel büyümeye yardımcı olmak, kentin refahını artırmak, sosyal hakkı gözetmek vb. vazifeler üstlenmiş durumdadır. Ülkemizde nüfusun artmasıyla işsizlik, yoksulluk vb. sorunların getirdiği bir sonuç olarak belediyeler, sosyal yardım ve sosyal hizmet alanında da önemli bir görev üstlenmektedir. Sosyal yardım ve hizmet alanında bir misyon üstlenen belediyeleri göz önüne alarak ve günümüzdeki kamu kuruluşlarının sayısal ortamdaki bilgilerinin rafine edilip faydalı bilgiler haline dönüştürülememesinden yola çıkılarak Malatya ilçe belediyesi olan Yeşilyurt Belediyesi örneği üzerinden bir tez çalışması gerçekleştirilmek istenmiştir. Bu amaçla Yeşilyurt Belediyesi Sosyal Yardım İşleri Müdürlüğü'nden veriler toplanmış ve bu verilere veri madenciliği yöntemleri uygulanarak ortaya çıkan anlamlı bilgilerle, bir Karar Destek Mekanizması oluşturmak amaçlamıştır.

2. VERİ MADENCİLİĞİ

Bilişim alanındaki gelişmelere bağlı olarak bilgisayarların kapasiteleri büyük oranda artmış ve buna bağlı olarak da büyük miktarlarda dijital veriler depolanmaya başlanmıştır. Bu büyük veriler içerisinde gizli kalmış ve değeri olan bilgileri elde etme işi basitçe veri madenciliği olarak tanımlanabilir [1]. Diğer bir deyişle bir kurum, kuruluş veya iş yerinde değerli olan herhangi bir bilgiyi açığa çıkarma ve bulma olarak tanımlanabilmektedir. Veri madenciliği aynı zamanda Veri Tabanlarında Bilgi Keşfi (VBK-Knowledge Discovery in Databases-KDD) olarak bilinir ve veri tabanlarındaki verilerden önemli olanları seçip çıkararak yararlı bilgiler meydana getirmektedir [1]. Veri madenciliği iş, bankacılık, sigorta, elektronik ve ticaret gibi alanlarda kullanılmaktadır.

Basitçe veri madenciliği çok büyük miktardaki verilerden en değerli verileri elde etmenin yoludur [17]. Veri madenciliği anlamlı kalıp ve kuralları bulabilmek için büyük miktardaki verilerden otomatik veya yarı otomatik bir şekilde analiz edip açıklayabilmektedir. Veri madenciliği bir kurumda belirli yöntemler kullanılarak üretilen tüm verilerin mevcut veya ortaya çıkabilecek gizli bilgileri seçip çıkarma süreci olarak kabul edilmektedir. Aslında veri madenciliği klasik istatistiksel yöntemlere benzemektedir [18]. Klasik istatistiksel yöntemler, iyi organize edilmiş ve çoğunlukla özetlenmiş verilerden oluşmaktadır. Fakat veri madenciliği milyonlarca hatta milyarlarca veriden daha fazlasıyla ilgilenmektedir.

2.1 Veriyi Bilgiye Dönüştürme

Günümüzde bilgi teknolojisi değişiklikleri altında sadece bilgisayar teknolojisi değil veri iletişim teknolojileri de bulunmaktadır [16]. Teknolojik gelişmelere bağlı olarak neredeyse bilginin her parçası dijital ortamlara kaydedilmektedir. Ancak kaydedilen bu veriler hakkında aşılması gereken üç önemli soru bulunmaktadır;

- İşletmelerde kaydedilen verilerin önemi nedir?
- Kaydedilmiş verilerin avantajları ne olabilir?
- Kaydedilmiş verilerden nasıl bilgi edinilebilir?

Çok farklı tiplerde toplanan veriler hakkında bazı analizler yaparak istatistiksel ve matematiksel yöntemler uygulamak, yeni kavram veri tabanı ve yeni analitik metotlar oluşturmak ve verileri yönlendirerek bir veri ambarı kullanmak, analiz etmek ve değerli veriler elde etmek gerekmektedir. Bu nedenle verinin bilgiye dönüştürülmesi büyük önem arz etmektedir.

2.2 Uygulama Alanları

Veri madenciliği alanlarında birçok yaygın kullanımlar bulunmaktadır. Örneğin pazarlama, bankacılık, sigortacılık ve e-ticaret alanları gibi alanlarda yaygın olarak kullanılmaktadır [3].

2.2.1 Pazarlama

Müşterinin satın alırken davranış biçimlerinin belirlenmesi ve demografik özellikte bağlantılarının ortaya çıkarılması, alışveriş sepeti analizinin yapılması, yönetim müşteri ilişkilerinin belirlenmesi, müşteri değerlendirilmesi ve satış tahmini (forecast) gibi pazarlama alanları veri madenciliğinde kullanılmaktadır.

2.2.2 Bankacılık

Farklı finansal göstergeler arasındaki gizli ilişkilerin ortaya konulması, kredi kartı dolandırıcılığının ve sahteciliğinin belirlenmesi, kredi kartı ödemelerine dayalı müşteri gruplarının belirlenmesi, kredi (credit) talebinin değerlendirilmesi gibi bankacılık alanları veri madenciliğinde kullanılmaktadır.

2.2.3 Sigorta

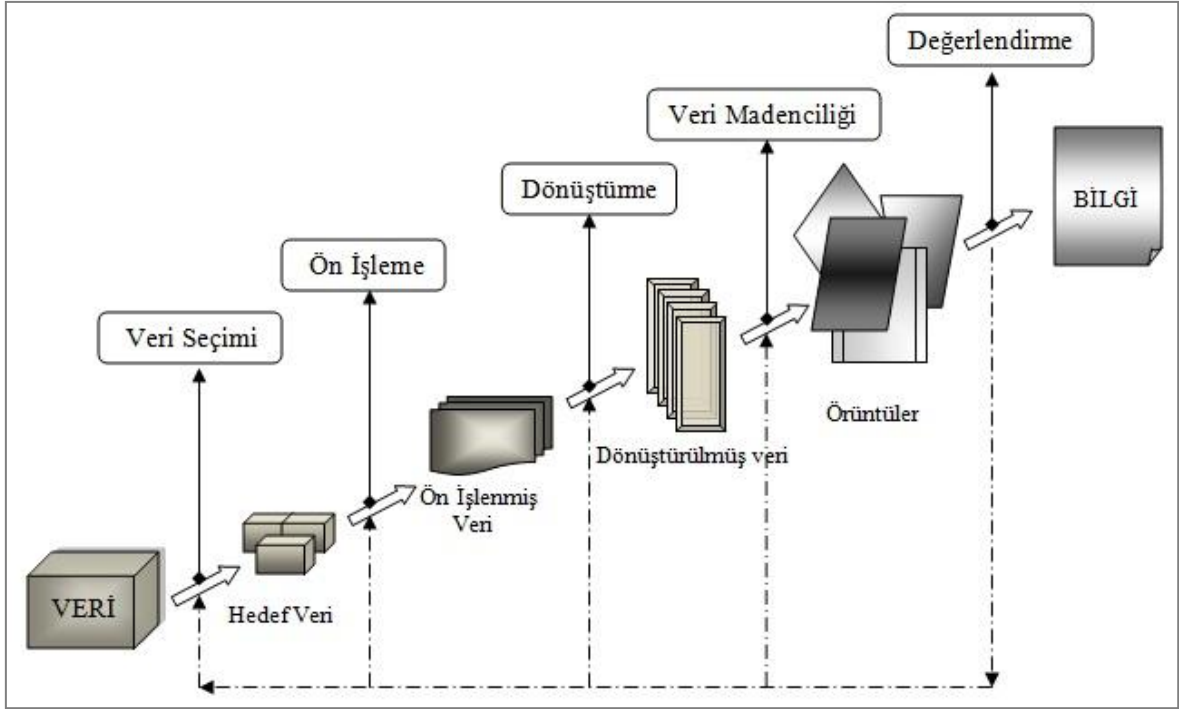
Yeni politikalar satın alacak tahminlerine potansiyel müşterilerin tahmini, hileli davranışların tespit edilmesi, riskli müşterilerin davranış biçimlerinin tespit edilmesi vb. sigorta alanları veri madenciliğinde kullanılmaktadır.

2.2.4 Elektronik Ticaret

Saldırının analiz edilmesi, E-CRM uygulamaları yönetimi, web sayfalarına ziyaretlerin analizi vb. elektronik ticaret alanları veri madenciliğinde kullanılmaktadır.

2.3 Veri Madenciliği Süreci

Veri madenciliği bir süreç olarak ele alındığında, verilerin temizlenmesi, farklı kaynaklardan alınan verilerin birleştirilmesi, verilerin boyutunun azaltılması, verilerin dönüştürülmesi, veri madenciliği algoritmalarının uygulanması, sonuçların sunumu ve değerlendirilmesi olmak üzere 6 aşamadan oluşmaktadır. Bu aşamalar Şekil 2.1’de verilmektedir.



Şekil 2.1. Veri madenciliği süreci [19]

2.3.1 Veri Temizleme

Veri tabanında meydana gelen tutarsız veriler gürültü olarak tanımlanmaktadır. Veri tabanında toplanan verilerden gürültülü veya ilgisiz olan verilerin çıkarılması veri temizleme olarak adlandırılır [1]. Bazı uygulamalarda, analiz edilecek bulguların istenen karakteristiğe sahip olmadığı görülmektedir. Örneğin, eksik ve uygun olmayan veri gibi tutarsız veri biçimleriyle karşılaşılabilir. Bu durumda eksik değerler yerine yeni değerlerin atanması işlemi gerekmektedir. Bu amaçla aşağıdaki belirtilen yöntemler uygulanabilir;

- Veri kümesinden eksik değerleri olan kayıtlar atılabilir.
- Eksik değerlerin yerine genel bir sabit değer yani tüm eksik değerlerin yerine aynı değer kullanılabilir.
- Tüm değerlerin ortalaması alınarak eksik değer tamamlanabilir.
- Her bir sınıfa ait değişkenlerin ortalamaları alınarak o sınıfa ait eksik değer tamamlanabilir.
- Eksik değerlerin yerine uygun bir tahmin modeli oluşturularak eksik değer tahmin edilebilir.

2.3.2 Veri Bütünleştirme

Veri bütünleştirme, farklı büyük veri tabanlarının birleştirildiği ve ayrıştırıldığı bir bilgi süreci olarak tanımlanmaktadır [20]. Bir veri türü içinde farklı veri tabanları toplanarak

verilerin dönüşümüne veri bütünleştirme denir [3]. Veri ambarı daha önceden tasarlanmışsa, veri madenciliği tekniği uygulamak için veri tümleştirme işlemi yapılmış olması gerekmektedir. Ancak bir veri ambarı daha önceden tasarlanmamışsa, veri madenciliği doğrudan yaklaşımla bu verilere uygulanmaktadır.

2.3.3 Veri İndirgeme

Veri indirgeme, aynı zamanda veri madenciliği problemi olarak bilinir, bu problem genellikle bilgi almak için bir ön-işleme yolu olarak görülmektedir [21]. Veri madenciliği uygulamalarının bazen analizi uzun zaman almaktadır. Sonuç analiz edildikten sonra değişmeyeceği düşünülüyorsa, veri veya değişken miktarı azaltılmaktadır. Veriyi indirmek için birçok yol vardır; Veri birleştirme veya veri küpü, Boyut küçültme, Veri sıkıştırma, Örnekleme ve Genelleme.

2.3.4 Veri Dönüştürme

Veriyi bazı durumlarda veri madenciliği analizine olduğu gibi eklemek doğru olmayabilir. Değişkenlerin ortalamaları ve değişkenlere ait varyanslar birbirinden önemli ölçüde farklı ise, en büyük ortalama ve varyanslara sahip değişkenlerin diğer değişkenler üzerindeki baskısı daha fazla olur ve bu durum rolleri önemli ölçüde azaltır. Ayrıca değişkenlerin çok büyük ve çok küçük değerleri analizin sağlam bir şekilde yapılmasını engeller. Böylece değişkenleri normalize veya standardize etmek için bir dönüşüm yöntemi uygulamak uygun bir yol olmaktadır [3].

2.3.4.1 Min-Max Normalizasyonu

Min-max normalizasyon yöntemi verileri 0 ile 1 arasında sayısal değere dönüştürmek için uygulanır. Bu yöntem, elde edilen veri içinde en büyük ve en küçük sayısal değer belirlenerek diğerlerini uygun yapıya dönüştürme prensibi üzerine dayanmaktadır. Yani dönüşüm bağıntısı olarak Denklem (2.1)'de ifade edilmektedir:

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2.1)$$

Bu bağıntıda, X^* dönüştürülen değerleri, X gözlem değerleri, $X_{\min}$ en küçük gözlem değeri ve $X_{\max}$ en büyük gözlem değeri olarak ifade edilmektedir.

2.3.4.2 Z-Score Standartlaştırma

İstatistik analizlerde yaygın olarak kullanılan başka bir dönüşüm biçimi z-score olarak adlandırılmaktadır. Bu yöntem, verilere ait ortalama ve standart sapmaya göre, bu verilerin yeni değerlere dönüşümünü temel almaktadır. Yani dönüşüm bağıntısı olarak Denklem (2.2)'deki gibi ifade edilmektedir:

$$X^* = \frac{X - \bar{X}}{\sigma_x} \quad (2.2)$$

Bu bağıntıda, X^* dönüştürülen değerleri, X gözlem değerleri, $\bar{X}$ verilerin aritmetik ortalaması ve σ_x gözlem değerlerinin standart sapması olarak ifade edilmektedir.

2.3.5 Veri Madenciliği Algoritmasının Uygulanması

Söz konusu süreçte, veri madenciliği yöntemi uygulamadan önce ilk gözlemlenen daha önce belirtilen işlemlerden en iyi sonuç verenin uygulanması işlemi olmaktadır. Veri hazır olduktan sonra ise veri madenciliği algoritmalarıyla ilgili konular uygulanmaktadır.

2.3.6 Sonuçları Sunum ve Değerlendirme

Verilere veri madenciliği algoritması uygulandıktan sonra, düzenlenen sonuçlar ilgili yerlere sunulur. Sonuçlar sıklıkla grafikler ile desteklenir. Örneğin, eğer bir hiyerarşik kümeleme modeli uygulanırsa, sonuçlar dendogram denilen özel bir grafik ile verilmektedir.

2.4 Veri Madenciliği Yöntemleri

Veri madenciliğinde tahmin en önemli hedeflerden biridir, bu yüzden veri madenciliğinde bu hedefe ulaşmak için çok sayıda yöntem ve algoritma geliştirilmiştir [22]. Bu yöntemlerin çoğu, otomatik öğrenme, örüntü tanıma ve istatistik tabanlı, denenmiş ve test edilmiş teknikler/yaklaşımlar olmaktadır. Veri madenciliği yöntemleri temel olarak üç ana başlıkta gruplandırılabilir; bunlar: Sınıflandırma, Kümeleme ve Birliktelik kuralıdır.

2.4.1 Sınıflandırma

Veri madenciliğinin sınıflandırılması veri tabanlarında gizli örüntülerin ortaya çıkarılması için sıklıkla kullanılan bir yöntemdir [3]. Veri sınıflandırması için özgün bir süreç izlenir. İlk olarak, mevcut veri tabanının bir kısmının eğitim amaçlı olarak kullanılması sınıflandırma kuralının oluşturulmasına yardımcı olur. Sonrasında, bu kural yeni bir durum ortaya çıktığında nasıl karar verileceğini belirlemeye yardımcı olmaktadır.

2.4.2 Kümeleme

Kümeleme, birbirleri arasındaki benzerlikleri göz önünde bulundurarak veri kümelerinin gruplandırılması işlemidir. Kümeleme yöntemi müşteri davranış analizi, hedeflenmiş pazarlama ve diğer pek çok analizler gibi çok farklı uygulamalarda geniş çapta kullanılmaktadır [23]. Böylece bu yöntem pek çok alanda uygulanabilir. Örneğin, pazarlama araştırmasında geniş ölçüde kullanılmaktadır. Ayrıca, örüntü tanımlama, görüntü işleme ve konumsal harita verilerinin analizinde de kullanılmaktadır.

2.4.3 Birliktelik Kuralı

Veri madenciliği yöntemlerinden Birliktelik kuralı, minimum destek ve güven eşiğini bulmak üzere öğelerin sıklığını veya tekrarlanmasını araştırmaktadır [4, 24]. Birliktelik Kuralı Yöntemi, bir veri tabanında bulunan verilerin ilişkilerini analiz etmek için ilişkili ve birlikte çalışılabilen durumları belirleyen bir veri madenciliği yöntemidir. Bu ilişkilerin belirlenmesi, birliktelik kurallarının elde edilmesine imkân sağlamaktadır. Birliktelik kuralları özellikle pazarlama alanında yaygınca kullanılmaktadır. Pazar Sepeti Analizi olarak adlandırılan uygulamalar bu tür veri madenciliği yöntemlerine dayanmaktadır [3]. Pazar Sepeti Analizi yardımıyla, bir müşterinin herhangi bir ürünü alıp almayacağını tahmin etmek ve bir ürünün satın alınabilme ihtimalini tespit etmek üzere bir dizi yöntemler uygulanmaktadır.

Örneğin Tablo 2.1’de verilen veriler göz önüne alındığında, müşteriler tarafından satın alınan veya alışveriş arabalarına eklenen ürünler arasındaki ilişkileri analiz etmek üzere birliktelik kuralı kullanılmaktadır.

Tablo 2.1. Beş (5) Müşterinin sepetindeki ürünler

Müşteri	Alışveriş arabasındaki ürünler
1	Makarna, Yağ, Su, Elma
2	Makarna, Ketçap
3	Ketçap, Yağ, Su, Alkol
4	Makarna, Ketçap, Yağ, Su
5	Makarna, Ketçap, Yağ, Alkol

Yukarıdaki verilere göre birliktelik analizi yapıldığında ve Apriori algoritması uygulandığında aşağıdaki sonuçlar elde edilmiştir;

{Ketap, Su}---> {Yağ}	(destek=0.4, güven=1.00)
{Ketap, Yağ}---> {Meyve suyu}	(destek=0.4, güven=0.67)
{Yağ, Su}---> {Ketap}	(destek=0.4, güven=0.67)
{Su}---> {Ketap, Yap}	(destek=0.4, güven=0.67)
{Yağ}---> {Ketap, Su}	(destek=0.4, güven=0.50)
{Ketap}---> {Yağ, Su}	(destek=0.4, güven=0.50)



3. SINIFLANDIRMA

Veri madenciliğinde sınıflandırılma veri tabanlarında gizli örüntülerin ortaya çıkarılması için sıklıkla kullanılan bir yöntemdir [3]. Veri sınıflandırması için özgün bir süreç izlenir. İlk olarak, mevcut veri tabanının bir kısmının eğitim amaçlı olarak kullanılması sınıflandırma kuralının oluşturulmasına yardımcı olur. Sonrasında, bu kural yeni bir durum ortaya çıktığında nasıl karar verileceğini belirlemeye yardımcı olmaktadır.

3.1 Model Başarım Ölçütleri

Bir modelin başarımını ölçerken birçok kavram kullanılmaktadır. Doğruluk, duyarlılık, hata-oranı ve F-ölçütü bunlardan en çok kullanılanlarıdır. Bu başarıım ölçümleri doğru ve yanlış sınıf olarak belirlenen sınıflara atanan örnek sayılarının sayısal değerleriyle yani başka bir ifadeyle modelin sınıflandırma işlemini ne kadar doğru yaptığıyla ilgilenmektedir. Bahsedilen modelin test işleminin gerçekleştirilmesiyle elde edilen sonuçların başarımı karışıklık matrisi (confusion matrix) ile gösterilmektedir. Karışıklık matrisi, bir modelden elde edilen doğru ve yanlış sınıflandırmaları gösteren bir tablo yapısını ifade etmektedir. Karışıklık matrisinde, test kümesindeki örneklerin gerçek sayıları tablonun satır kısmında, kolon kısmında ise modelin tahmin değerleri gösterilmektedir. Tablo 3.1’de gösterimi bulunmaktadır [25-27].

Tablo 3.1. İki sınıf için karışıklık matrisi

		Tahmin Edilen Sınıf	
		Pozitif	Negatif
Gerçek Sınıf	Pozitif	Doğru Pozitif (DP)	Yanlış Negatif (YN)
	Negatif	Yanlış Pozitif (YP)	Doğru Negatif (DN)

Tablo 3.1’de DP ile ifade edilen kısım sınıflandırıcının modelde pozitif sınıfını doğru tahmin ettiği nesne sayısıdır. YP ile ifade edilen kısım sınıflandırıcının model üzerinde pozitif olarak tahmin ettiği ama negatif sınıfa ait olan nesne sayısıdır. YN ifadesi, sınıflandırıcının modelde negatif olarak tahmin edip pozitif sınıfa ait olan nesne sayısıdır.

DN ile ifade edilen kısım ise sınıflandırıcının negatif olarak doğru tahmin yürüttüğü nesne sayısıdır. Karışıklık matrisi sınıf sayısı ne kadar artarsa o kadar büyük boyutlara ulaşmaktadır. Tahmin edilen sınıf ve gerçek sınıf için tanımlanan birkaç standart terim bulunmaktadır [26-29].

Duyarlılık veya Doğru Pozitif Oranı (DPO): Doğru bir şekilde tanımlanan pozitif durumların oranı:

$$Duyarlilik = DP / (DP + YN) \quad (3.1)$$

Doğru Negatif Oranı (DNO): Doğru bir şekilde sınıflandırılmış negatif durumların oranı:

$$DNO = DN / (DN + YP) \quad (3.2)$$

Kesinlik: Doğru olan pozitif durumların tahmin edilme oranı:

$$Kesinlik = DP / (DP + YP) \quad (3.3)$$

Yanlış Negatif Oranı (YNO): Negatif olarak yanlış sınıflandırılmış pozitif durumların oranı:

$$YNO = YN / (YN + DP) \quad (3.4)$$

Yanlış Pozitif Oranı (YPO): Pozitif olarak yanlış sınıflandırılmış negatif durumların oranı

$$YPO = YP / (YP + DN) \quad (3.5)$$

Doğruluk: Doğru olan tahminlerin toplam sayıya oranı:

$$Dogruluk = (DP + DN) / (DP + DN + YP + YN) \quad (3.6)$$

Hata Oranı: Negatif ve pozitif olarak yanlış sınıflandırılmış durumların toplam durumlara oranı:

$$Hata Oranı = (DP + YN) / (DP + DN + YP + YN) \quad (3.7)$$

F-Ölçütü: Kesinlik ve Duyarlılık ölçütlerinin birlikte değerlendirilerek anlamlı sonuçlar verilmesi amacıyla F-Ölçütü tanımlanmıştır. F-Ölçütü bu iki ölçütün harmonik ortalamasıdır.

$$F - \text{Ölçütü} = (2 \times Duyarlilik \times Kesinlik) / (Duyarlilik + Kesinlik) \quad (3.8)$$

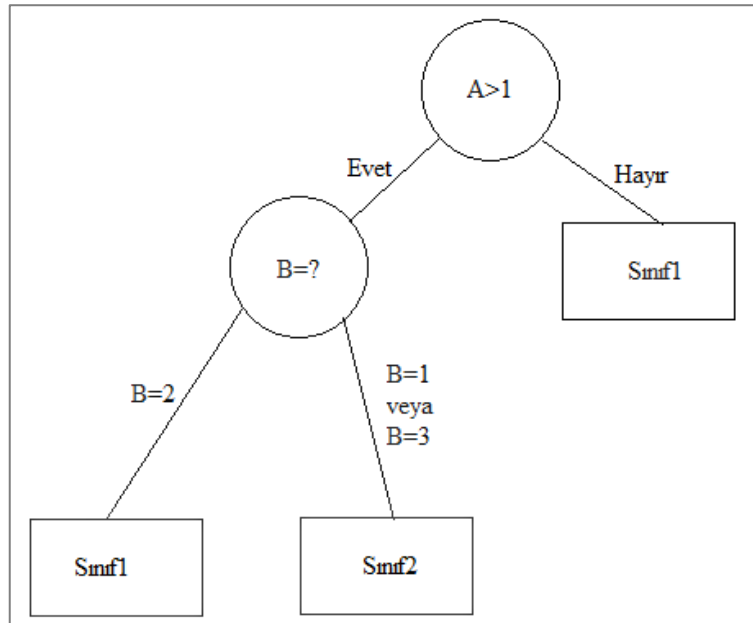
3.2 Sınıflandırma Algoritmaları

Bu bölümde, Veri Madenciliği yöntemlerinden sınıflandırma algoritmaları hakkında incelemeler yapılmıştır. Bu yöntemlere ait birkaç algoritma sunulmuştur. Sınıflandırma algoritmalarından Karar Ağacı ile Sınıflandırma, ID3, C4.5, J48, K-En Yakın Komşu

(KNN), Sade Bayes Sınıflandırıcı (BS), Yerine Koyarak Örnekleme (YKÖ), Rastgele Orman algoritmaları (ROS), Sıralı Minimal Optimizasyon (SMO) ve Yerel Ağırlıklı Öğrenme (YAÖ) algoritmaları açıklanmıştır.

3.2.1 Karar Ağacı ile Sınıflandırma

Bir karar ağacı, regresyon ve sınıflandırmadaki soruları veya sorunları çözmek üzere kullanılan her türlü etkili malzeme veya araç olarak tanımlanabilir [30]. Sınıflandırmanın görevi, verilen işaretlenmemiş bir nokta için işaret veya sınıfı tahmin etmektir. Sınıflandırıcı bir özelliğin eğitim kümesi ve değerlerine (sınıf işaretleri) bağlı olarak veriyi sınıflandırır (bir model oluşturur) ve bunu yeni verilerin sınıflandırılmasında kullanır. Sınıflandırma, bir veri kümesindeki özniteliklerin gözlemlerinin benzerliklerine bağlı olarak bir veri kümesindeki gruplaşmaları oluşturur. Hedef, önceden görülmemeyen kayıtlara mümkün olduğunca doğru bir şekilde bir sınıf atanmasının gerçekleştirilmesidir. Makine öğrenmesi adı altında, uygulamalı istatistikte pek çok karar algoritmaları geliştirilmiştir [3]. Karar ağacı sınıflandırması çoğunlukla anlaşılması kolay ve uygulaması ekonomik (ucuz) olduğu için kullanılmaktadır [31]. Karar ağaçları sınıflandırma algoritmasının uygulanması için altyapı iyi olmalıdır. A ve B olarak iki girdi olarak alınırsa, $A > 1$ ve $B = 2$ sınıf1'de, $B = 1$ ve $B = 3$ sınıf2'de bulunmaktadır. B 'nin değerine bakmadan $A \leq 1$ şartına uyan örnekler sınıf1'de bulunur. Şekil 3.1'de gösterimi bulunmaktadır.



Şekil 3.1. X ve Y nitelikleri üzerine uygulanan teste dayanan bir basit karar ağacı [3].

Karar Ağacında Dallanma Kriterleri;

Karar ağaçlarının en önemli unsurlarından birisi de akışın kökten dallara veya uç yapraklara tanımlanmasıdır. Aşağıdaki algoritmalar şu şekilde kategorize edilebilir [3, 32]:

- Entropi tabanlı algoritma (ID3, C4.5)
- Sınıflandırma ve regresyon ağaçları
- Hafıza tabanlı sınıflandırma algoritmaları (KNN)

3.2.2 ID3 Algoritması

ID3 Algoritması, Ross Quinlan tarafından keşfedilmiş, veri kümesinden bir karar ağacı oluşturmak için kullanılan bir algoritmadır [33]. ID3 tipik olarak makine öğrenmesi ve doğal dil işleme alanlarında kullanılmaktadır. Karar ağacı tekniği, sınıflandırma işlemini modellemek için bir karar ağacı oluşturulmasını içeren bir algoritmadır. Ağaç oluşturulduktan sonra, veri tabanındaki her bir değişken grubuna uygulanır ve o değişken grubu için sınıflandırmayı oluşturur.

Entropi

Entropi bir sistemin belirsizliğinin bir ölçütü olarak tanımlanabilir; sistem ya belirsizlik ya da ihmal durumundadır [34]. Genel olarak, daha büyük düzensizlik daha büyük entropi anlamına gelir. Verilen bir S örnek kümesinde, T sınıfı kullanılarak S , S_1 ve S_2 olarak iki aralığa bölümlendirilirse, bölümlendirme sonrası bilgi kazancı;

$$I(S, T) = \frac{|S_1|}{|S|} H(S_1) + \frac{|S_2|}{|S|} H(S_2) \quad (3.9)$$

Entropi küme içindeki örneklerin sınıf dağılımına göre hesaplanır. Verilen m sınıf için S_1 entropisi;

$$H(S_1) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (3.10)$$

p_i , S_1 'de i sınıfının olasılığıdır.

Karar Ağacında Entropi

Eğitim için veri tabanından elde edilen kayıt kümesi göz önünde bulundurulduğunda, sınıf özelliği eğitim kümesinden elde edilen değerlere göre $\{C_1, C_2 \dots C_k\}$ k sınıfa

bölünmüştür. Bu sınıf ile ilgili olarak ortalama miktarda bilgi gerekli olabilir. T sınıf değerlerinde bulunan küme için P_T sınıfı olasılık dağılımının hesaplanması şöyledir [3];

$$P_T = \left(\frac{|C1|}{|T|}, \frac{|C2|}{|T|}, \dots, \frac{|Ck|}{|T|} \right) \quad (3.11)$$

$$H(T) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (3.12)$$

Karar Ağacında Dallanma Kriterlerinin Seçimi

X nitelik değerine bağlı olan T 'nin $T_1, T_2, \dots, T_n$ olarak ayrılması göz önünde bulundurulduğunda T öğelerinin sınıfı Denklem (3.13)'teki gibi belirlenebilir.

$$H(X, T) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i) \quad (3.13)$$

Bilgi T veri tabanının X testi için bölümlendirilmesi ile elde edilir. Bu ifadeye Kazanç ölçümü denir. Hesaplama Denklem (3.14)'teki gibidir;

$$\text{Kazanç}(X, T) = H(T) - H(X, T) \quad (3.14)$$

3.2.3 C4.5 Algoritması

ID3 algoritmasının dezavantajlarını azaltmak veya ortadan kaldırmak için bir uzantı olarak C4.5, karar ağaçlarını oluşturmak üzere kullanılan, iyi bilinen bir algoritmadır [35]. C4.5 algoritması ile oluşturulan karar ağaçları sınıflandırma için kullanılabilir ve bu nedenle C4.5 aynı zamanda istatistiksel sınıflandırıcı olarak anılmaktadır [3]. C4.5 algoritması ID3 algoritmasını geliştirmek üzere bir dizi değişiklik yapmıştır.

Nümerik Değerleri İşleme

Eğitim örneklerinde ilk olarak uygun eşik değeri belirlemek için niteliğin aldığı değerler sıralanır. Bu yüzden nitelik değerleri $\{v_1, v_2, \dots, v_m\}$ biçimini alır ve v_i ve $v_i + 1$ aralığının orta noktası alınarak t_i eşik değeri olarak belirlenebilir [3]:

$$t_i = \frac{v_i + v_{i+1}}{2} \quad (3.15)$$

C4.5'te, $\{v_i, v_i + 1\}$ aralığının en küçük olan değeri, eşik değeri olarak alınır.

Bilinmeyen Nitelik Değerleri

C4.5'te, bilinmeyen değerlerin bilinen değerlerin göreceli sıklığına göre olasılıklı bir şekilde dağıtılması olarak tanımlanabilmektedir. Yeni kazanç kriteri Denklem (3.17)'deki gibi olacaktır:

$$F = \frac{\text{verilen özellik için veritabanında bilinen değerli örnek sayısı}}{\text{veri kümesindeki örneklerin toplam sayısı}} \quad (3.16)$$

$$\text{Yeni Kazanç}(X) = F(H(T) - H(X, T)) \quad (3.17)$$

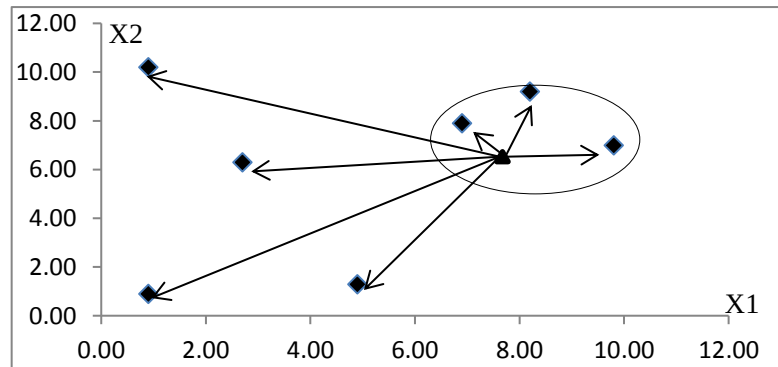
3.2.4 J48 Algoritması

J48, J. Ross Quinlan tarafından geliştirilen popüler bir karar ağacı algoritmasıdır. C4.5 algoritması temeline dayanan bu algoritma, Weka veri madenciliği aracında açık kaynak kodlu olarak C4.5 algoritması için yazılmış bir algoritma olarak bilinmektedir. Böl ve yönet şeklinde yukarıdan aşağıya doğru bir ağaç yapısı oluşturur. Bu algoritmada öznelilikler düğüm noktası meydana getirecek şekilde yerleştirilir ve eğitim verisine göre yapraklar oluşturulur [36-39].

3.2.5 K-En Yakın Komşu Algoritması

Bir bellek tabanlı sınıflandırma algoritması olan K-en yakın komşu sınıflandırma algoritması, en temel ve basit sınıflandırma yöntemlerinden birisi olarak tanımlanmaktadır. Bu sınıflandırma algoritması, veri dağılımı hakkında çok az veya hiç ön bilgi olmadığında sınıflandırma araştırması için ilk tercihlerden biri olarak düşünülebilmektedir [3]. K-en yakın-komşu sınıflandırıcısı genel olarak test örneği ve belirtilen eğitim örnekleri arasındaki öklid mesafesine dayanmaktadır. x_i , p özellikli $(x_{i1}, x_{i2}, \dots, x_{ip})$, giriş örneği olsun, n giriş örneklerinin toplam sayısı ($i=1, 2, \dots, n$) olsun ve p de özelliklerin toplam sayısı ($j=1, 2, \dots, p$) olsun. Örnek x_i ve x_j arasındaki Öklid mesafesi Denklem (3.18)'deki gibi tanımlanmaktadır:

$$d(i,j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3.18)$$



Şekil 3.2. Gözlemlenen $\blacktriangle$ noktasına en yakın $k=3$ komşunun belirlenmesi [3]

K-En yakın komşu algoritmasında, gözlem değerinden oluşturulan her türlü küme aşağıdaki işlemlere göre gerçekleştirilmelidir:

- a) K parametresinin belirlenmesi
- b) Mesafenin hesaplanması
- c) Minimum mesafenin belirlenmesi
- d) Seçili satırlardan ilgili sınıfların belirlenmesi
- e) Yeni gözlem sınıfının oluşturulması

Ağırlıklı Oylama

K-en yakın komşu algoritması sadece k adet komşuyu göz önüne almasından dolayı her zaman uygun olmayabilmektedir. Bu nedenle sadece k adet komşuyu seçmek yerine bunların ağırlıklı oylamaları kullanılması daha uygun sonuçlar verebilmektedir.

Ağırlıklı mesafenin hesaplaması Denklem (3.19)'da görüldüğü gibi bir ilişkiye dayanır [3,40];

$$d(i,j)' = \frac{1}{d(i,j)^2} \quad (3.19)$$

Burada $d(i, j)$ i ve j arasındaki Öklid mesafesi anlamına gelmektedir. Mesafe her bir sınıf değeri için hesaplanır ve ağırlıklı oylama değerini elde etmek için toplanır. En yüksek ağırlığa sahip oylama sonucu yeni gözlem sınıfı olarak kabul edilir.

3.2.6 Bayes Sınıflandırıcısı

İstatistiksel sınıflandırma modeli olan Bayes sınıflandırıcı, Bayes teoremini kullanan bir algoritmadır. Bir örneğin hangi olasılıkla hangi sınıfa ait olduğunu kestirir [41].

Örneğin: bir torbada bazı cisimlerin olduğu varsayarsak, kırmızı yuvarlak cisimler olduğunu varsayıldığında, varsayım "bu cisim bir elma" olacaktır. Burada $P(H)$ "önsel olasılık" olacaktır. Yani başlangıçta bu olasılığın ne olduğu biliniyor. Fakat $P(X | H)$ olasılığı H koşulu üzerine olduğundan, "sonrasal olasılık" olarak değerlendirilir, yani X kırmızı ve yuvarlak olarak biliniyorsa sonuç "bu bir elma" olacaktır. Bu yüzden Bayes ilişkisi şu şekildedir:

$$P(H / X) = \frac{P(X | H)P(H)}{P(X)} \quad (3.20)$$

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(X)} \quad (3.21)$$

$P(X | C_i)$ hesaplama işleminde ağırlığı azaltmak için olasılık birleştirilebilir. Bunu gerçekleştirmek için örnekteki X_i değerlerinin bağımsız elde edildiği düşünülür;

$$P(X / C_i) = \prod_{k=1}^n P(x_k | C_i) \quad (3.22)$$

Bilinmeyen X örneğinin sınıflandırması için, (3.21)'da $P(X | C_i)$ içinde ki paydalar eşit olduğundan pay değeri ile karşılaştırma yapılmalıdır. Bu değerlerden en büyüğü seçilmeli ve bilinmeyen örneğin bu sınıfa ait olduğu söylenmelidir.

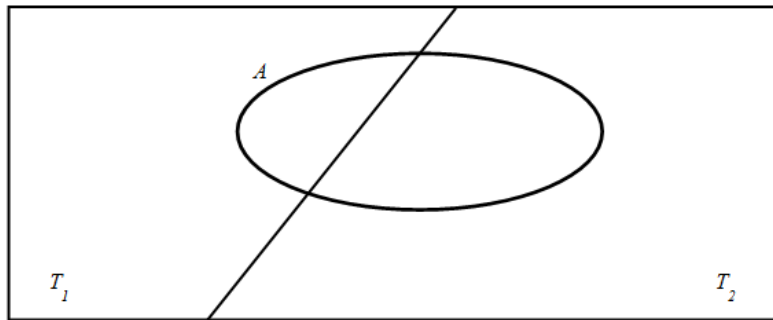
$$\operatorname{argmax}_{C_i} \{P(X | C_i)P(C_i)\} \quad (3.23)$$

Yukarıda kullanılan sonrasal olasılık ifadesi, en büyük sonrasal sınıflandırma yöntemi (*Maximum A Posteriori classification = MAP*) olarak da bilinir. Bu durumda, sonuç olarak Bayes sınıflandırıcısı için aşağıdaki bağıntı kullanılabilir:

$$C_{MAP} = \operatorname{argmax}_{C_i} \prod_{k=1}^n P(x_k | C_i) \quad (3.24)$$

Bayes Teoremi

Olasılık hesabında çok önemli bir yere sahip olan bu teorem kullanılarak sınıflandırma işlemi gerçekleştirmek mümkün olmaktadır. T_1 ve T_2 şeklinde iki olayı göz önüne aldığımızda bu iki olayın bağdaşmadığını $T_1 \cap T_2 = \emptyset$ olarak gösterimi aşağıda açıklanmaktadır.



Şekil 3.3. T_1 , T_2 ve A olayları

Bir A olayını T_1 ve T_2 cinsinden alındığında bu koşullu olasılık $P(T_1 | A)$;

$$P(T_1 | A) = \frac{P(A | T_1)}{P(A)} \quad (3.25)$$

Şekil 3.3'te görüldüğü gibi A olayı hem T_1 hem de T_2 olayında meydana gelmiştir. A için aşağıdaki ilişki şu şekilde yazılabilir:

$$A = (T_1 \cap A) \cup (T_2 \cap A)$$

Yani A olayının olasılığı şu şekilde yazılabilir;

$$P(A) = P(T_1 \cap A) + P(T_2 \cap A)$$

Aşağıdaki ilişkinin sonucunu elde etmek için Denklem (3.26)'da ki formül kullanılabilir;

$$P(T_1 | A) = \frac{P(A | T_1)}{P(T_1 \cap A) + P(T_2 \cap A)} \quad (3.26)$$

Eğer olay $T_1, T_2, \dots, T_n$ kümesinin bir grubu ise ve olasılık sıfıra eşit değilse, A olayının olay T_j de meydana gelme olasılığı Denklem (3.27)'deki şekilde hesaplanır;

$$P(T_j | A) = \frac{P(A | T_j)}{\sum_{i=1}^n P(A \cap T_i)} \quad (3.27)$$

Bu şekilde, $P(A | T_j) = P(A \cap T_j) / P(T_j)$ Denklem (3.28)'de ki gibi elde edilir:

$$P(T_j | A) = \frac{P(A \cap T_j)}{\sum_{i=1}^n P(A \cap T_i)} \quad (3.28)$$

Koşullu Olasılık

Koşullu olasılık, bazı koşulların olup olmadığına bağlı olarak bir olayın meydana gelmesi şeklinde ifade edilebilir. Başka bir deyişle, bir olayın koşullu olasılığı meydana gelmiş diğer bir olayın ilave bilgisi ile elde edilmiş bir olasılıktır [42]. İki uyumlu (bağdaşan) olay olarak A ve B göz önünde bulundurulduğunda, A ve B olayları arasında ortak noktalar bulunur. Bu durumda $A \cap B \neq \emptyset$ denebilir. Bir B olayının meydana gelmesi tamamen A olayına bağlıdır. $P(B|A)$ olasılığı Denklem (3.29)'da ki gibidir;

$$P(B | A) = \frac{P(A \cap B)}{P(A)} \quad (3.29)$$

Verilen bilgilere göre, bileşik olasılık bağıntısı Denklem (3.30)'da ki şekilde bulunur;

$$P(A \cap B) = P(A)P(B | A) \quad (3.30)$$

A ve B gibi iki olayın ardı ardına meydana gelme olasılığı iki olayın olasılıklarının çarpımına eşit olmaktadır ve şu şekilde yazılabilir;

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Şu şekilde de açıklanabilir;

$$P(A \cap B) = P(B)P(A | B)$$

Eğer yerine yeniden yazılırsa Denklem (3.31)'de ki bağıntı ile elde edilebilmektedir;

$$P(B | A) = \frac{P(B)(A | B)}{P(A)} \quad (3.31)$$

Bayes Sınıflandırıcılarında Sıfır Değer Sorunu

Bayes sınıflandırıcı kullanıldığında herhangi bir olasılık değeri sıfır ise tüm sonuç sıfır olacağından sınıflandırma problemi doğru sonuç vermeyebilir. Yani diğer olasılığın çıkarımı sıfırdır ve sonuç bir anlam ifade etmez. Bu sorunu çözmek için, k gibi küçük bir değer her orana eklenebilir. Bu yüzden, $\frac{n}{d}$ olasılığı için Denklem (3.32) kullanılır:

$$\frac{n + kp}{d + k} \quad (3.32)$$

Burada k , 0 ve 1 arasında bir değerdir. Çoğunlukla 1 tercih edilir. p değeri 0.5 olarak alınır.

3.2.7 Yerel Ağırlıklı Öğrenme Algoritması

Yerel Ağırlıklı Öğrenme Algoritması (Locally weighted learning-LWL). Aynı zamanda hafıza tabanlı öğrenme, örnek tabanlı öğrenme, tembel öğrenme ve çekirdek yoğunluk tahmini, benzerlik araştırması olarak da bilinir ve esnek bir yapıya sahiptir [43]. LWL, robotik sistemlerin özerk uyarlamalı kontrolü sırasında karmaşık olguları öğrenmek için yararlı temsiller ve eğitim algoritmaları sağlayan istatistiksel öğrenme tekniklerinden bir sınıftır [41].

Yerel Ağırlıklı Öğrenme, örnek tabanlı bir algoritma olup örneklerle ağırlıklar atayarak bu ağırlıklar işlenmekte ve algoritma bu şekilde çalışmaktadır. Bu ağırlık, tahmin edilecek noktanın göre giriş değişken alanındaki eğitim noktasının konumuna bağlıdır. Tahmin noktasına yakın eğitim gözlemleri genellikle daha yüksek ağırlık alır. Bu ağırlık, tahmin edilecek noktanın göre giriş değişken alanındaki eğitim noktasının konumuna bağlıdır [44].

LWL yöntemleri parametrik değildir ve mevcut tahmini, verilerin yalnızca bir alt kümesini kullanan yerel işlevlerle yapılır. LWL'nin arkasındaki temel fikir, tüm işlev alanı için küresel bir model oluşturmak yerine, her bir ilgi noktası için, sorgu noktasının komşu verisine dayalı olarak bir yerel model oluşturulmasıdır. Bu amaçla, her veri noktası, veri noktasının tahmini etkisini ifade eden bir ağırlıklandırma faktörü haline gelir. Genel olarak, mevcut sorgu noktasına yakın olan veri noktaları, uzaktaki veri noktalarından daha yüksek bir ağırlık almaktadır. LWL'ye yeni eğitim noktaları eklemek kolaydır ve bu yaklaşım LWL'yi doğru bir fonksiyon yaklaşım metodu yapar [45]. Sade Bayes algoritmasına benzer bir şekilde sınıflandırma, Lineer Regresyon gibi de regresyon yapabilen bir algoritmadır [46].

3.2.8 Sıralı Minimal Optimizasyon Algoritması

Sıralı Minimal Optimizasyon (SMO) Algoritması, destek vektör makinesi tabanlı bir algoritma olup John Platt tarafından geliştirilmiştir. Destek vektör makinelerinin eğitim süreci boyunca genellikle optimizasyon problemlerinin çözümünde kullanılır. SMO tüm eksik değerleri değiştirir ve kategorik(nominal) nitelikleri ikili niteliklere dönüştürür. Varsayılan olarak tüm öz nitelikleri normalleştirir [47, 48].

Destek Vektör Makineleri(Support Vector Machine-SVM) elle yazılmış karakter tanıma, yüz tanıma, yaya algılama ve metin kategorileştirme gibi çok çeşitli problemlerde deneysel olarak iyi bir genelleme performansı gösterir. Bununla birlikte, SVM'lerin kullanımı halen küçük bir araştırmacı grubuyla sınırlıdır. Bunun bir nedeni, SVM'ler için özellikle büyük problemlerde eğitim algoritmalarının yavaş olmasıdır. Başka bir nedeni ise, SVM eğitim algoritmalarının karmaşık, çözümü zor ve ortalama bir mühendis tarafından uygulanmasının güç olmasıdır.

Kavramsal olarak basit, uygulanması kolay yeni bir SVM öğrenme algoritması açıklanmaktadır. Yeni SVM algoritması genel olarak daha hızlı ve zor SVM sorunları için standart SVM eğitim algoritmasından daha iyi ölçeklendirme özelliklerine sahiptir ve bu algoritmaya Sıralı Minimum Optimizasyon (SMO) adı verilir. SMO, bir iç döngü olarak sayısal kuadratik programlamayı (quadratic programming, QP) kullanan önceki SVM öğrenme algoritmaları yerine analitik bir QP adımı kullanır. Sıralı Minimum Optimizasyon (SMO), herhangi bir ekstra matris saklama olmadan ve sayısal QP optimizasyon adımlarını hiç kullanmadan SVM QP problemini çabucak çözebilecek basit bir algoritmadır. SMO, yakınsamayı sağlamak için Osuna teoremi kullanılarak genel QP sorununun QP alt

problemlerine ayrılmasını sağlar. Sıralı Minimum Optimizasyon (SMO), herhangi bir ekstra matris saklama olmadan ve sayısal QP optimizasyon adımlarını hiç kullanmadan SVM QP problemini çabucak çözebilecek basit bir algoritmadır. SMO, yakınsamayı sağlamak için Osuna teoremi kullanılarak genel QP sorununun QP alt problemlerine ayrılmasını sağlar. Önceki yöntemlerin aksine, SMO her adımda mümkün olan en küçük optimizasyon problemini çözmeyi seçmektedir. Standart SVM QP problemi için olası en küçük optimizasyon problemi iki Lagrange çarpanı içerir, çünkü Lagrange çarpanları lineer eşitlik kısıtına uymalıdır. SMO, her aşamada, ortaklaşa optimize etmek için iki Lagrange çarpanı seçer, bu çarpanlar için en uygun değerleri bulur ve SVM'yi yeni en uygun değerleri yansıtacak şekilde günceller. SMO'nun avantajı, iki Lagrange çarpanı için analitik olarak çözüm bulabilmesidir. Böylece sayısal QP optimizasyonu tamamen önlenir. Algoritmanın iç döngüsü, bütün bir QP kitaplığı rutin olarak çağırmak yerine kısa bir C kodu cinsinden ifade edilebilir. Algoritma sürecinde daha fazla optimizasyon alt problem çözülsede, her alt problemin çok hızlı gerçekleşmesiyle genel QP problemi hızla çözülür. Buna ek olarak, SMO hiç ekstra matris saklama gerektirmez. Bu nedenle, çok büyük SVM eğitim problemleri sıradan bir kişisel bilgisayar belleğinin içine sığabilir. SMO'da hiçbir matris algoritması kullanılmadığı için, sayısal tahmin problemlerine daha az duyarlıdır. SMO için iki bileşen vardır: iki Lagrange çarpanı için bir çözümleme yöntemi ve çarpanları optimize etmek için bir buluşsal yöntemdir [49-51].

3.2.9 Yerine Koyarak Örneklemeye Algoritması

1994 yılında Leo Breiman tarafından önerilen bir algoritma olan Bagging (Bootstrap aggregating) algoritması, yerine koyarak örneklemeye olarak adlandırılır [52, 53]. Bu algoritma, temel öğreniciyi tekrar eğitmek amacıyla bir eğitim setinden yeni eğitim setleri üreten bir yöntemdir. Bagging Algoritmasında belirlenen n adet örneğin bulunduğu eğitim setinden aynı şekilde n örnekli bir eğitim seti yerine koymalı rastgele seçimle üretilmektedir [52]. Başka bir ifadeyle Bagging algoritması, bütün örneklerin belirli bir yüzdesini alarak oluşturduğu n adet örnekle temel öğreniciyi eğitmekte ve eğitilen her bir temel öğrenici için bir sınıflandırma işlemi gerçekleştirmektedir. Bahsedilen örnekler veri kümesindeki örneklere birebir benzememektedir. N adet bir standart eğitim seti D olarak alındığında, Bagging algoritması örneklemeye ve değiştirme yöntemi kullanarak yeni D_i eğitim setleri üretir, her bir eğitim setinin büyüklüğü n 'dir [49]. Bu durumda birtakım eğitim örnekleri yeni eğitim kümesinde bulunmazken bazıları ise birden çok kez bulunmaktadır.

Topluluktaki her bir temel öğrenici, birbirinden farklı örnekler içeren eğitim kümeleriyle bu şekilde eğitilirler ve öğrenicinin kararları birleştirilir. Böylece test kümesi üzerinde daha başarılı sonuçlar alınabilmektedir [54, 55].

3.2.10 Rastgele Orman Sınıflandırma Algoritması

Kolektif bir sınıflandırma algoritması olan Rastgele Orman (RO) Breiman ve Cutler tarafından geliştirilmiştir [52]. Bagging yöntemi ile Ho tarafından 1998 yılında önerilen The Random Subspace adlı teknikle bir araya getirilerek yeni bir yöntem oluşturulmuştur [56]. Rastgele Orman Sınıflandırma algoritması sınıflandırma işlemi esnasında birden çok karar ağacı kullanarak sınıflandırma değerinin bulunmasını hedeflemektedir [57, 58]. Topluluk öğrenme yöntemi olan Rastgele Orman Sınıflandırma algoritması, birbirinden ayrı olarak oluşturulmuş sınıflama ve regresyon karar ağaçlarıyla karar ormanı topluluğunu meydana getirmektedir. Rastgele Orman sınıflandırma algoritması, yalnız bir karar ağacı oluşturmak yerine birden fazla sayıda ve birden fazla değişkenli ağaçların her birinin ayrı eğitim kümeleriyle eğitiminden meydana gelen kararların birleşimini önermektedir. Bir sınıflandırıcı kullanmak yerine birden fazla sınıflandırıcı oluşturmakta ve daha sonra üretimi gerçekleştirilen sınıflandırıcıların tahmininden alınan oylar aracılığıyla yeni veriyi sınıflandırabilen bir öğrenme algoritması olmaktadır [57, 58]. Karar ormanının meydana gelişi sırasında elde edilen sonuçların toplanıp birleştirilmesiyle en son tahminin oluşturulması hedeflenmektedir [56]. Karar ağacına benzer bir sisteme sahip olan Rastgele Orman, temel farkı eğitim sürecinde birden çok ağacın üretilmesine izin vermesidir. RO yönteminde, ormanı meydana getiren karar ağaçları Bootstrap metodu kullanılarak seçilen farklı örneklerden toplanır. Bu yöntem için orijinal veri seti kullanılır [52]. Bilinen makine öğrenme yöntemleri içerisinde RO yöntemi, eşsiz bir tahmin geçerliliği ve model yorumlama olanağı sağlamaktadır. Ayrıca topluluk yöntemlerdeki tekniklerin iyileştirilmiş özelliklerini içermesi sebebiyle RO yöntemi daha iyi genellemeler sunmakta ve geçerli tahminlerde bulunmaktadır [56].

4. SOSYAL YARDIM VERİ KÜMESİ İLE SINIFLANDIRMA UYGULAMASI

Bu bölümde, Yeşilyurt Belediyesi Sosyal Yardım İşlerinden veriler alınmış ve bu veri setine veri madenciliği algoritmaları uygulanarak en iyi sonucu veren algoritma bulunmaya çalışılmıştır. Toplanan veriler üzerine veri madenciliği algoritmalarından sınıflandırma algoritmaları uygulanmıştır. Sınıflandırma algoritması uygulanırken veriler eğitim ve test kümesi olarak iki kümeden oluşmaktadır. Eğitim seti üzerinde eğitilen veri kümeleri sınıflandırma modellerini oluşturur. Oluşturulan modeller test setinde test edilir ve doğrulukları hesaplanır. Bu şekilde sosyal yardım alan kişiler için eğitim ve test seti olarak iki küme kullanılmıştır. Sınıflandırma Algoritmaları üzerinde deneme yapılarak en yüksek doğruluğa sahip algoritma belirlenmiştir.

4.1 Veri Kümesi

Yeşilyurt Belediyesi Sosyal Yardım İşlerinden 190 adet anket verisi alınmıştır. Alınan anket verileri incelenip analiz edildiğinde anket üzerinde bulunan bazı kısımların işaretlenmediği tespit edilmiştir. Örneğin anket üzerinde “Evde İhtiyaç Olan Eşya” diye bir alan bulunup bu alan yetkililer tarafından boş bırakılmıştır. Boş bırakılıp doldurulmayan kısımlar dikkate alınmayıp gerekli alanlar üzerinden işlem yapılmıştır. 190 veri içerisinde 20 tanesi yalan beyan ya da adres değişikliği sebebiyle yardım iptali yapılan kişilerden oluşmuştur. Bu sebeplerin bir sonucu olarak, anketten elde edilen veriler öncelikle ön işlemden geçerek düzenlenmiş ve gürültülü verilerden arındırıldıktan sonra geriye kalan 170 adet veri üzerinde çalışma gerçekleştirilmiştir. Tablo 4.1’de, anket sorularının yanı sıra her soruya ait görülen frekans ve yüzde değerleri de gösterilmiştir. Tablodan görüldüğü üzere bu verilerin 94 tanesi değerlendirme sonucunda yardım almış kişilerden, 76 tanesi ise yardım alamamış kişilerden oluşmuştur.

Tablo 4.1. Anket soruları ve cevapları

1. Ailenin Net Geliri	Frekans	Yüzde
Net asgari ücretin %50'sine kadar geliri olanlar	111	65,29
Net asgari ücretin %50'sinden fazla olup,%100'üne kadar geliri olanlar	38	22,35
Net asgari ücretin %100'ünden fazla olup %160'a kadar geliri olanlar	21	12,35
2. Evin Durumu	Frekans	Yüzde
Kendi Evi (Brüt 70 m2 den fazla ise)	12	7,06
Kendi Evi(Brüt 70 m2 den küçük ise)	74	43,53
Kirada	53	31,18
Kirası karşılanıyor	21	12,35
3. Evin Isınma Durumu	Frekans	Yüzde
Kalorifer	71	41,76
Sobalı	99	58,24
4. Ailenin Kişi Sayısı	Frekans	Yüzde
Bir	12	7,06
İki	29	17,06
Üç	32	18,82
Dört	35	20,59
Dörtten fazla	62	36,47
5. Ailede Okuyan Sayısı(ilkokul, ortaokul, lise)	Frekans	Yüzde
Yok	81	47,65
Bir	32	18,82
İki	32	18,82
Üç	21	12,35
Dört	2	1,18
Dörtten fazla	2	1,18
6. Ailede Üniversite Okuyan Sayısı	Frekans	Yüzde
Bir	151	88,82
İki	18	10,59
Üç	1	0,59
7. Ailede yaşlılık maaşı alan var mı?	Frekans	Yüzde
Evet	10	5,88
Hayır	160	94,12
8. Ailede özürlülük maaşı alan var mı?	Frekans	Yüzde
Evet	33	19,41
Hayır	137	80,59
9. Yardım alınan bir kurum var mı?	Frekans	Yüzde
Yok	81	47,65
Kaymakamlık veya diğer kamu kurumu	77	45,29
Vakıf, dernek vb. yerlerden alınan yardım	12	7,06
10. Ailede çalışabilir kişi sayısı?	Frekans	Yüzde
Çalışabilir kişi yok	28	16,47
Bir kişi	74	43,53
Birden fazla	68	40,00
11. Evin Fiziki Durumu	Frekans	Yüzde
Kötü	12	7,06
Normal	143	84,12
İyi	15	8,82
12. Ailede özürlü bakım ücreti alan var mı?	Frekans	Yüzde
Evet	12	7,06
Hayır	158	92,94
13. Yardım Sonucu	Frekans	Yüzde
Evet	94	55,29
Hayır	76	44,71
TOPLAM	170	100

Bu tez çalışmasında, veri madenciliği algoritmalarını uygulayabilmek için veri madenciliği araçlarından biri olan WEKA (Waikato Environment for Knowledge Analysis), yazılımı kullanılmıştır. WEKA, nesneye yönelik bir programlama dili olan java programlama diliyle yazılmış ve içerisinde birçok sınıflandırma, kümeleme ve birliktelik kuralları algoritmaları barındıran kullanımı kolay bir yazılımdır [59, 60]. WEKA, aracının sınıflandırma modülü kullanılarak veri setinin “yardım alma” niteliği üzerinde sınıflandırma işlemi gerçekleştirilmiştir.

4.2 Veri Kümesi Üzerinde Sınıflandırma Uygulamaları

Bu çalışmada, veri seti sınıflandırma işlemi için birçok sınıflandırma algoritması kullanılmış fakat en iyi sonucu veren 5 farklı sınıflandırma algoritması seçilerek elde edilen sonuçlar karşılaştırılmıştır. Çok sayıda sınıflandırma algoritması içeren WEKA yazılımından Bayes Sınıflandırıcı (BS), Yerine Koyarak Örneklem (YKÖ), Rastgele Orman Sınıflandırma (ROS), Sıralı Minimal Optimizasyon (SMO) ve Yerel Ağırlıklı öğrenme (YAO) algoritmaları bu veri seti için en iyi sonuçların elde edildiği algoritmalar olarak gözlemlenmiştir. 170 aileden toplanan veriler ön işleme alınıp WEKA aracının veri format şekline dönüştürülmüştür. Şekil 4.1’de arff formatının küçük bir gösterimi bulunmaktadır.

```
@relation yardım

@attribute ID REAL
@attribute Soru1 { 1,2,3 }
@attribute Soru2 { 1,2,3,4 }
@attribute Soru3 { 1,2 }
@attribute Soru4 { 1,2,3,4,5 }
@attribute Soru5 { 0,1,2,3,4,5 }
@attribute Soru6 { 0,1,2 }
@attribute Soru7 { 1,2 }
@attribute Soru8 { 1,2 }
@attribute Soru9 { 0,1,2 }
@attribute Soru10 { 1,2,3 }
@attribute Soru11 { 1,2,3 }
@attribute Soru12 { 1,2 }
@attribute sonuc { E,H }

@data

1,2,3,2,5,2,0,2,2,1,3,2,2,H
2,1,3,2,1,0,0,2,2,1,2,1,2,H
3,1,2,2,2,0,0,2,1,1,1,2,1,H
4,1,2,1,2,0,0,2,2,1,1,2,2,H
5,2,2,1,2,0,0,2,2,0,3,2,2,H
6,1,1,1,3,0,0,2,1,0,3,2,1,E
```

Şekil 4.1. Arff Formatı Gösterimi

4.3 Veri Kümesi Sınıflandırma Sonuçları

Tablo 4.1’de toplanan verilerin anket soruları ve aynı zamanda cevaplarının frekans ve yüzde değerleri bulunmaktadır. Veri seti 170 ailenin bilgisini içermektedir. Veri setinde ankete katılan ailelerin net geliri, ev durumu, evin ısınma durumu, ailenin kişi sayısı, evin fiziki durumu, ailedeki çalışabilir kişi sayısı gibi bilgiler bulunur ve sınıflandırma işlemi gerçekleştirilerek ailenin yardım alıp almayacağı konusunda tahmin yapılması amaçlanmıştır. Anketin yardım sonucunda kişiye yardım sağlandıysa “Evet”, sağlanmadıysa “Hayır” olarak ifade edilerek Evet cevabı bir sınıfı, Hayır cevabı ise diğer sınıfı oluşturacak şekilde 2 sınıflı bir sınıflandırma yapısı oluşturulmuştur. Verilerden elde edilen başarımlar oranları Tablo 4.2’de karşılaştırmalı olarak gösterilmektedir. Tablodan görüldüğü üzere bu 5 algoritma arasında en iyi sonuç Yerel Ağırlıklı Öğrenme algoritması ile elde edilmiştir.

Tablo 4.2. Sınıflandırma sonuçları

<i>Ağırlıklı Ort.</i>	YKÖ	ROS	BS	SMO	YAÖ
TP Oranı	80.6	79.4	79.4	78.8	81.8
FP Oranı	21.7	22.4	22.2	23.2	20.8
Kesinlik	81.2	79.6	79.5	79.1	82.7
F-Ölçütü	80.2	79.2	79.2	78.6	81.4
ROC Alanı	82.7	83.3	80.1	77.8	79.5
Başarımlar	80.6	79.4	79.4	78.8	81.8

Tablo 4.2’de verilen değerler “Evet” ve “Hayır” sınıfına ait ağırlıklı ortalama veriler olup en yüksek başarımlar değeri %81.8 ile Yerel Ağırlıklı Öğrenme (LWL) algoritmasından elde edilmiştir. Bu algoritmaya ait karışıklık matrisi Tablo 4.3’te verilmektedir. Tablodan görüldüğü üzere Evet sınıfında bulunan 94 verinin 87 tanesi doğru olarak tahmin edilebilmiştir. Bu sonuç Evet sınıfına ait başarımlar değerinin %92.55 olduğunu göstermektedir.

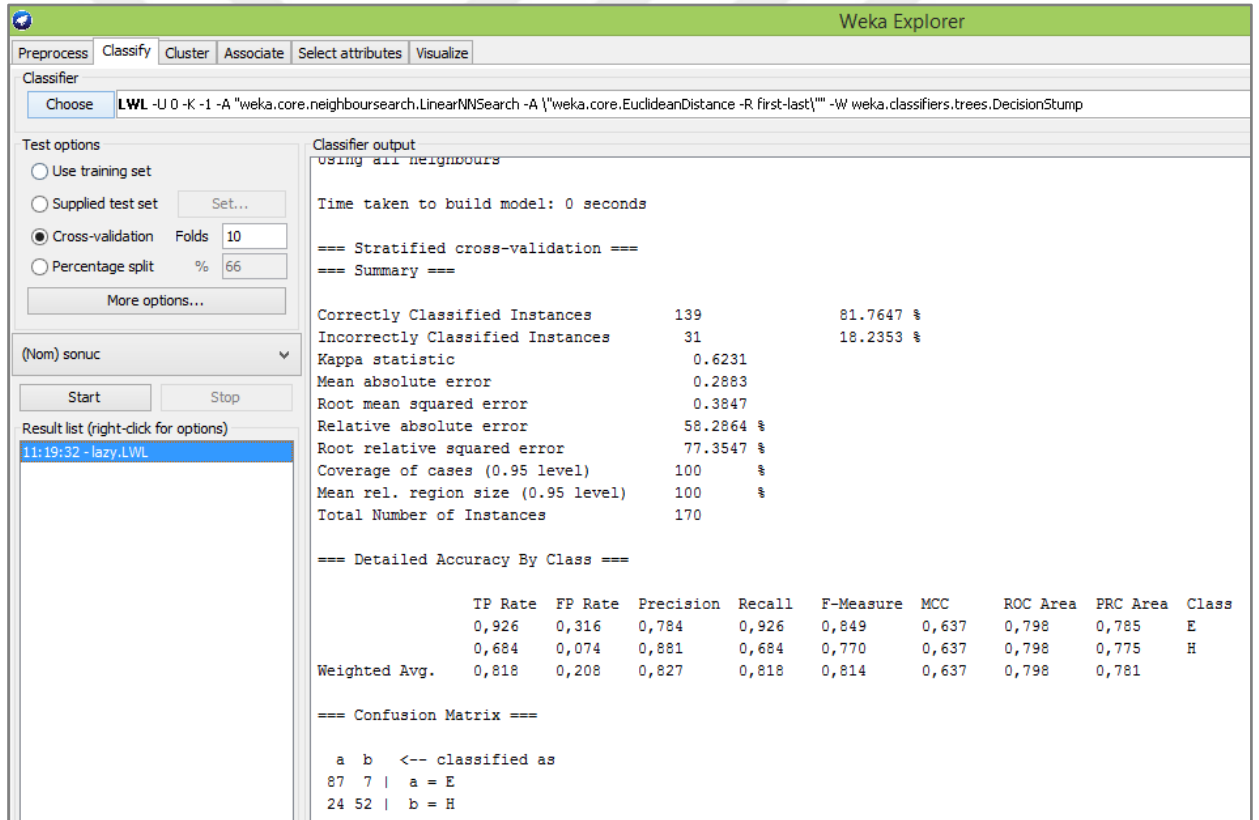
Tablo 4.3. YAÖ Algoritması karışıklık matrisi

	Tahmin		Sınıf
	Evet	Hayır	
	87	7	Evet
Gerçek	24	52	Hayır

Tablo 4.4. YAÖ Algoritması doğruluk çizelgesi

	İşlem	Sonuç
Duyarlılık	$87 / (87 + 7)$	0.93
Kesinlik	$87 / (87 + 24)$	0.78
Doğruluk	$(87 + 52) / (87 + 52 + 24 + 7)$	0.82
Hata-Oranı	$(24 + 7) / (87 + 52 + 24 + 7)$	0.18
F-Ölçütü	$(2 \times 0.93 \times 0.78) / (0.93 + 0.78)$	0.85

En iyi sonucu veren algoritma olarak Yerel Ağırlıklı Öğrenme (LWL) algoritmasının WEKA aracındaki gösterimi Şekil 4.2’de bulunmaktadır.



Şekil 4.2. Yerel Ağırlıklı Öğrenme Algoritması sonuçlarının WEKA aracıda gösterimi

5. GELİŞTİRİLEN KARAR DESTEK SİSTEMİ

Bu bölümde, Malatya Büyükşehir İlçe Belediyesi olan Yeşilyurt Belediyesi örneği üzerinde web tabanlı bir karar destek sistemi geliştirilmiştir. Bu karar destek sistemi geliştirilirken Bölüm 4'te tespit edilen en iyi üç sınıflandırıcıdan elde edilen kurallar kullanılmıştır. Bu üç sınıflandırıcı, Yerine Koyarak Öğrenme, Sade Bayes ve Yerel Ağırlıklı Öğrenme sınıflandırıcılarıdır. Bu sınıflandırma algoritmalarından elde edilen karar ve kural tabloları kullanılarak web ortamında gerekli bilgilerin alınıp yardım alma sonucunun geri dönebileceği bir karar destek sistemi yazılımı kodlanarak gerçekleştirilmiştir.

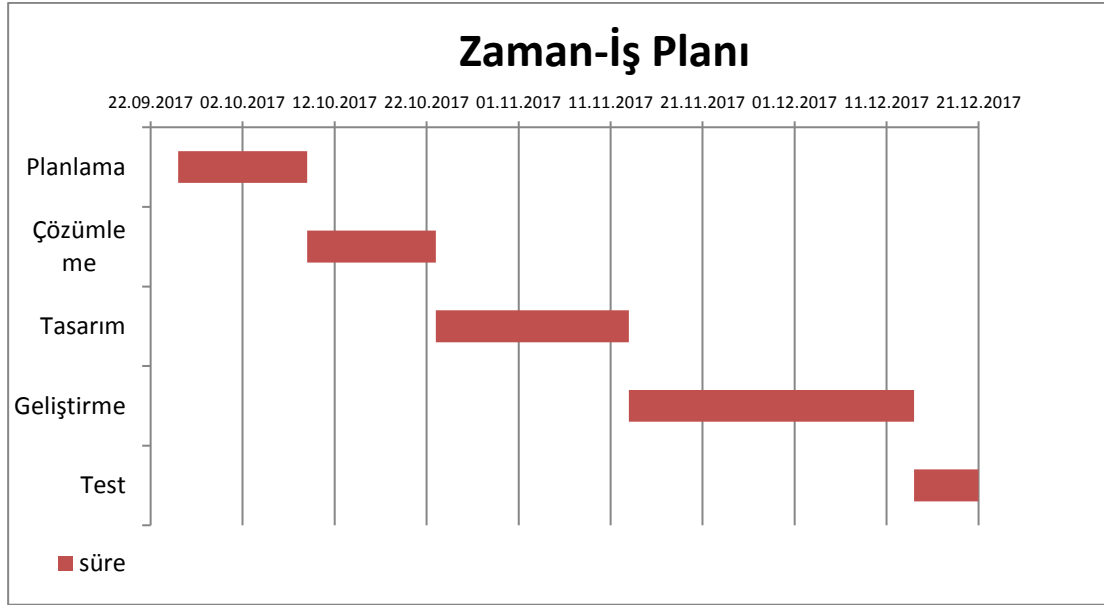
5.1 Karar Destek Sisteminin Amacı

Yeşilyurt Belediyesi örneği üzerinden bir Karar Destek Mekanizması geliştirmek amacıyla öncelikle Yeşilyurt Belediyesi Sosyal Yardım İşleri Müdürlüğü'nde yardım dağıtım işleyişi incelenmiştir. Bu işleyiş, ilgili kişiler tarafından yardım talebinde bulunan ailenin yaşadığı yer incelenip aile ile ilgili gerekli bilgilerin bir anket formuna doldurulması ve daha sonra toplanan bilgiler kurul tarafından incelenip ailenin yardım alıp almayacağına dair karara bağlanması şeklinde gerçekleşir. Belediyelerin sosyal yardım dağıtımını esnasında eldeki verileri hızlı bir şekilde kullanabilmesi ve ilgili komitede doğru karar verilmesi açısından web ortamında bilgilerin kaydedilerek ailenin yardım alıp almaması konusunda sonucun dönebileceği bir yazılım geliştirilmiştir.

5.2 Karar Destek Sisteminin Planlanması

Yazılım gerçekleştirilirken yazılım geliştirme süreçlerinden planlama kısmı ile başlanıp diğer aşamalarla sonuçlandırılmıştır. Planlama aşaması, diğer adımlar konusunda nasıl bir yol izleneceğini belirlemesi açısından önem arz eden bir aşamadır.

Genel olarak planlama sürecinde, Malatya Yeşilyurt Belediyesi Sosyal Yardım İşleri'nden alınan verilerin veri tabanına aktarılacak şekilde olması için veri tabanı tablo alanları ve özelliklerine dair isterler oluşturulmuş ve aynı zamanda kullanıcı girişi sağlanacağından yetkilendirme yapılabilecek tablo yapısı düşünülmüş ayrıca yazılım projesinin zaman-iş planı, teknik karmaşıklık tablosu vb. belirlenmiştir. Şekil 5.1 de karar destek sisteminin zaman-iş plan tablosu bulunmaktadır.



Şekil 5.1. Karar Destek Sistemi Zaman-İş Planı Tablosu

Tablo 5.1. Karar Destek Sistemi Genel Sistem Puanlama Tablosu

Genel Sistem Puanlama Tablosu	
Uygulama, güvenilir yedekleme ve kurtarma gereksinimi	4
Veri iletişimi	5
Performans kritikliği	4
Çevrim içi veri girişi gereksinimi	5
Bir ara işlem için birden çok ekran gereksinimi	0
Girdiler, çıktılar, kütükler ya da sorguların karmaşıklığı	2
İşsel işlemler karmaşıklığı	4
Tasarlanacak kodun yeniden kullanılabilirliği	4
Dönüştürme ve kurulumun tasarımıdaki önemi	5
Sistemin birden çok yerde yerleşik farklı kurumlar için geliştirilmesi	3
Tasarlanan uygulamanın kolay kullanılabilir ve kullanıcı tarafından kolayca değiştirilebilir olması	5

Tablo 5.1 de genel sistem puanlama tablosu bulunmaktadır. 0'dan 5'e kadar ağırlık değerleri verilmiştir;

- 0: hiçbir etkisi yok,
- 1: çok az etkisi var,
- 2: etkisi var,
- 3: ortalama etkisi var,

- 4: önemli etkisi var,
- 5: mutlaka olmalı.

Karar destek sisteminde, kullanıcı girdileri sisteme girişı saęlanan kullanıcıların sistemde yaptıkları işlemleri ifade ederken kullanıcı çıktısı, sistemi kullanan kullanıcının sistemde yaptığı işlem den aldığı sonuçtur. Sistemi kullanabilmek için oluşturulan veri tabanlarının sayısı kütük olarak adlandırılır. Kullanıcı sorgulama ise, admin yetkisinde olan kullanıcının yaptığı işlemlere denilir.

Kullanıcı Girdi Bilgisi

- Giriş yapacak kullanıcının kullanıcı adı ve şifre bilgisi
- Sosyal yardım formuna kaydı yapılacak aile bilgisi

Kullanıcı Çıktı Bilgisi

- Sosyal yardım formuna kaydı gerçekleşen ailenin özet form bilgisi
- Sosyal yardım formuna kaydı gerçekleşen ailenin yardım sonuç bilgisi

Kullanıcı Sorgu Bilgisi

- Kullanıcı giriş bilgileri sorgulama
- Aile yardım sonucu sorgulama

Kütük Bilgisi

- Menü kütüğü (Kullanıcı tabanlı yetkilendirme için)
- Aile Bilgisi Kütüğü
- Kullanıcı Kütüğü

Dışsal Arayüz Bilgisi

- Kullanıcı giriş ekranı
- Aile bilgilerinin kaydını gerçekleştirecek form ekranı
- Kaydı gerçekleşen aile bilgisinin özet form ekranı
- Ailenin yardım sonuç ekranı
- Kullanıcı ekleme ve yetkilendirme ekranı

5.3 Karar Destek Sisteminin Çözümlemesi

Çözümleme kısmında, yazılımın planlama aşamasındaki isterler gerçekleştirilmiş olup veri tabanı tabloları ve alanları oluşturulmuştur. Tablolar hem anket verilerinin kaydedilebileceği şekilde, hem de kullanıcı yetkilendirmesi ve şifre kontrolü yapılabilecek biçimde gerçekleştirilmiştir. Şekil 5.2’de admin kullanıcısının Şekil 5.3’te ise normal kullanıcının yapacağı işlemler gösterilmiştir.

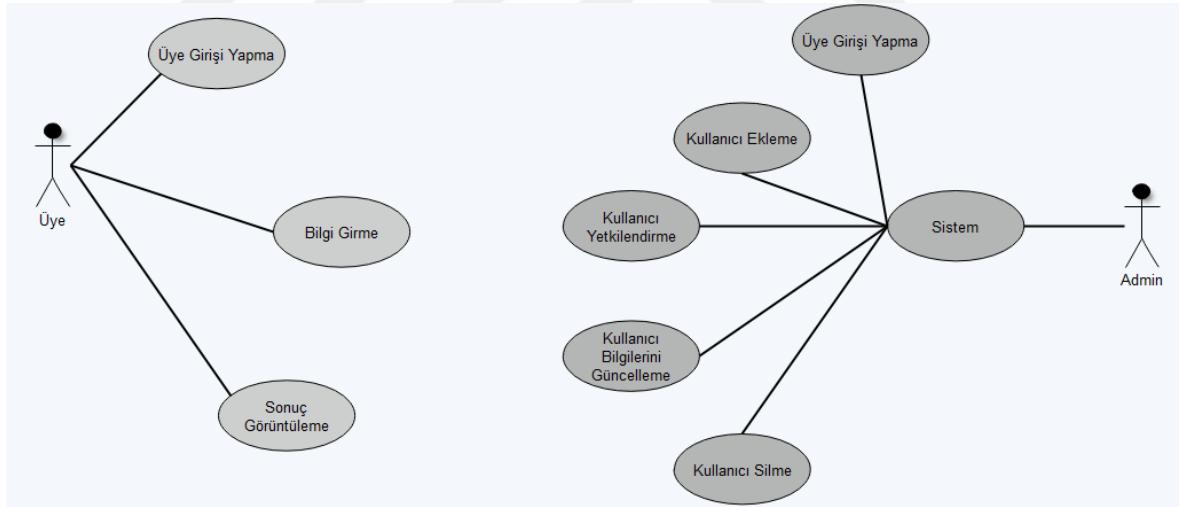
Üye Girişi Yapma	• Admin kullanıcı adı ve şifresi ile giriş yapabilecektir.
Kullanıcı Ekleme	• Admin kullanıcı ekleyebilecektir.
Kullanıcı Yetkilendirme	• Admin istediği kullanıcıya gerekli gördüğü yetkiyi verebilecektir.
Kullanıcı Bilgilerini Güncelleme	• Admin gerekli durumlarda kullanıcı bilgilerini güncelleyebilecektir.
Kullanıcı Silme	• Admin uygun olmayan kullanıcıları silebilecektir.

Şekil 5.2. Karar Destek Sisteminde Admin Kullanıcı İşlemleri

Üye Girişi Yapma	• Üyeler kullanıcı adı ve şifresi ile giriş yapabilecektir.
Bilgi Girme	• Üyeler aile bilgilerini girecek ve kaydedebilecektir.
Sonuç Görüntüleme	• Üyeler ailenin yardım sonucunu görüntüleyebilecektir.

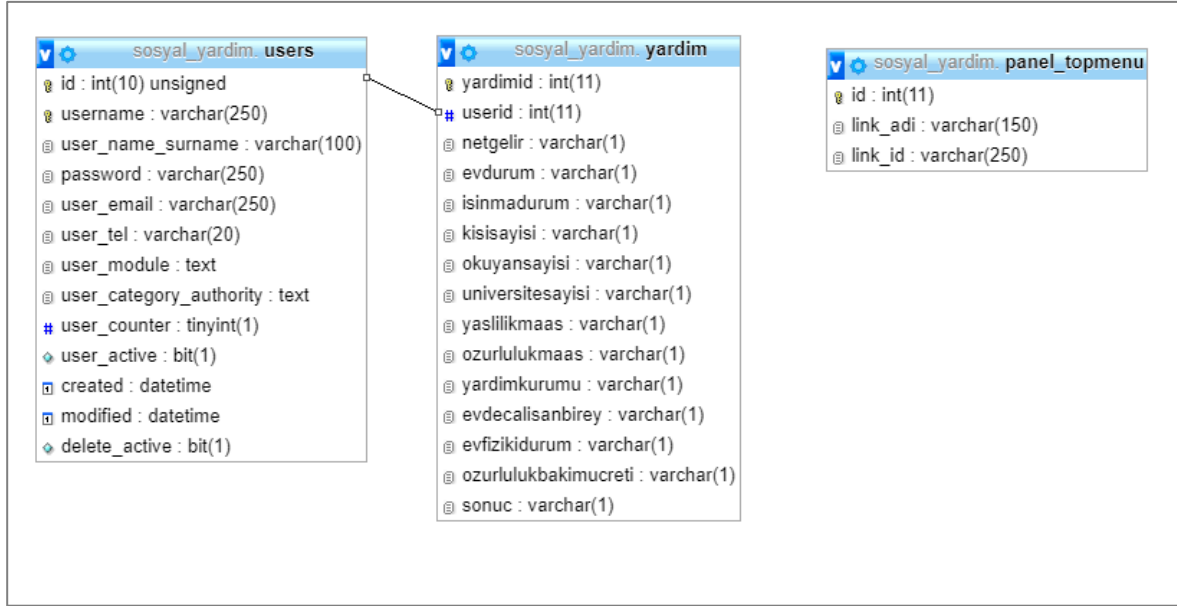
Şekil 5.3. Karar Destek Sisteminde Normal Kullanıcı İşlemleri

Şekil 5.2 ve 5.3'te açıklanan admin ve normal kullanıcı işlemlerinin Kullanım Durumu Diyagramındaki (Use Case Diagram) gösterimi Şekil 5.4'te bulunmaktadır.



Şekil 5.4. Karar Destek Sistemi Kullanım Durum Diyagramı

Şekil 5.5'te ise Karar destek sisteminin veri modeli bulunmaktadır. Şekilde görüldüğü gibi yardım sonucu verilecek ailenin bilgilerinin tutulduğu tablo “yardim” tablosu, giriş yapacak kullanıcı bilgisinin bulunduğu tablo “users” tablosu ve giriş yapan kullanıcının hangi menülere gireceğinin belirlenmesi ise “panel_topmenu” tablosunda tutulmaktadır.



Şekil 5.5. Karar Destek Sistemi Veri Modeli

5.4 Karar Destek Sistemini Tasarlama ve Geliştirme

Karar destek sistemi anket verisi girme, kullanıcı ekleme ve yetkilendirme şeklinde ekranlara sahip olup kullanıcının kolaylıkla giriş yapıp ailenin bilgilerini girebileceği bir görünüm tasarlanmıştır.

Geliştirme aşamasında ise planlama, tanımlama ve tasarım aşamaları bittikten sonra yazılımın kodlamasına geçilmiştir.

Yazılım, Şekil 5.6'da görüldüğü gibi bir kullanıcı giriş ekranına sahiptir. Kullanıcı adı ve şifreyle giriş sağlanmaktadır.



Şekil 5.6. Panel Giriş Ekranı

Panele giriş yapan kullanıcının karşısına Şekil 5.7’deki gibi bir form ekranı gelmektedir.

Sosyal İnceleme Tespit Tutanağı

Ailenin net geliri:

- ☐ Net asgari ücretin %50'sine kadar(%50 dahil) geliri olanlar.
- ☐ Net asgari ücretin %50'sinden fazla olup, %100'üne kadar(%100 dahil) geliri olanlar
- ☐ Net asgari ücretin %100'ünden fazla olup, %160'a kadar(%160 dahil) geliri olanlar

Evin Durumu:

- ☐ Kendi evi(Brüt 70 metrekareden fazla ise).
- ☐ Kendi evi(Brüt 70 metrekareden küçük ise)
- ☐ Kirada
- ☐ Kirası karşılanıyor

Evin Isınma Durumu

- ☐ Kalorifer
- ☐ Sobalı

Ailenin Kişi Sayısı:

- ☐ Bir
- ☐ İki
- ☐ Üç
- ☐ Dört
- ☐ Dörtten fazla

Şekil 5.7. Veri Girişi Ekranı

Şekil 5.7’da görüldüğü gibi “Veri Girişi” yazan ekrana gelen kullanıcı aile ile ilgili bilgileri girip kaydını yapabileceği bir form ekranıyla karşılaşmaktadır. Yazılım, anket üzerinden alınan aile bilgilerini bir form yardımıyla girilmesini sağlar ve veri tabanına kaydını gerçekleştirir. Şekil 5.8’de aile bilgilerinin girildiği form ekranının tümden görünümü bulunmaktadır.

Sosyal İnceleme Tespit Tutanağı

Ailenin net geliri:

- ☐ Net asgari ücretin %50'sine kadar(%50 dahil) geliri olanlar.
- ☐ Net asgari ücretin %50'sinden fazla olup, %100'üne kadar(%100 dahil) geliri olanlar
- ☐ Net asgari ücretin %100'ünden fazla olup, %160'a kadar(%160 dahil) geliri olanlar

Evin Durumu:

- ☐ Kendi evi(Brüt 70 metrekareden fazla ise).
- ☐ Kendi evi(Brüt 70 metrekareden küçük ise)
- ☐ Kırada
- ☐ Kirası karşılanıyor

Evin Isınma Durumu

- ☐ Kalorifer
- ☐ Sobalı

Ailenin Kişi Sayısı:

- ☐ Bir
- ☐ İki
- ☐ Üç
- ☐ Dört
- ☐ Dörtten fazla

Aile Okuyan Sayısı (İlkokul,ortaokul ve lise):

- ☐ Yok
- ☐ Bir
- ☐ İki
- ☐ Üç
- ☐ Dört
- ☐ Dörtten fazla

Ailede Üniversite Okuyan Sayısı:

- ☐ Bir
- ☐ İki
- ☐ Üç

Ailede Yaşlılık Maaşı Alan Var Mı?

- ☐ Evet
- ☐ Hayır

Ailede Özürürlük Maaşı Alan Var Mı?

- ☐ Evet
- ☐ Hayır

Yardım Alınan bir kurum var mı?

- ☐ Yok
- ☐ Kaymakamlık ve diğer kamu kurumu
- ☐ Vakıf, dernek vb. yerlerden alınan yardım

Ailede Çalışabilecek Birey Durumu Sorgulayan Kriterler

- ☐ Çalışabilir kişi yok
- ☐ Çalışabilir 1 kişi var
- ☐ Çalışabilir 1 den fazla kişi var

Evin Fiziki Durumu

- ☐ Kötü
- ☐ Orta
- ☐ Normal

Ailede Özürürlük Bakım Ücreti Alan Var Mı?

- ☐ Evet
- ☐ Hayır

Kaydet

Şekil 5.8. Aile Bilgileri Giriş Ekranı

Şekil 5.8'deki ekran, Yeşilyurt Belediyesi Sosyal Yardım İşleri Müdürlüğü'nün kullandığı form yapısına göre tasarlanarak bilgilerin girilmesi sağlanmaktadır. Şekil 5.8'de görüldüğü gibi ailenin gelir durumu, ev durumu, ısınma durumu, kişi sayısı, çalışabilir kişi sayısı gibi bilgiler bulunup kaydı gerçekleştirilmektedir. Ailenin kaydı gerçekleştirildikten sonra Şekil 5.9'daki gibi bir "Özet Form" adı altında özet bilgiler gelmektedir.

37

ÖZET FORM

- Ailenin gelir düzeyi düşük durumdadır.
- Aile kirada oturmaktadır.
- Aile kaloriferli bir evde oturmaktadır.
- 3 kişilik bir ailedir.
- Ailede 0 tane ilkokul, lise vb. okuyan sayısı vardır.
- Ailede 0 tane üniversite okuyan sayısı vardır.
- Evde devlet kurumunun sağladığı yaşlı bakım vb. maaş türlerinden herhangi birinden alınan bir gelir bulunmamaktadır.
- Aile herhangi bir kurumdan yardım almamaktadır.
- Ailede çalışabilir durumda kimse bulunmamaktadır.
- Evin fiziki hali kötü durumdadır.

Yerel Ağırlıklı Öğrenme

Yerine Koyarak Örnekleme

Sade Bayes Sınıflandırıcı

Şekil 5.9. Özet Form Ekranı

Özet bilgiler bulunan formda, Bölüm 4’te bahsedilen veri setinden en iyi sonucu veren üç algoritmanın hesaplama yapacağı butonlar bulunmaktadır. Belirtilen butonlar seçildiğinde, Yerel Ağırlıklı Öğrenme, Yerine Koyarak Örnekleme, Sade Bayes Sınıflandırıcı algoritmalarından oluşturduğumuz karar yapısına göre if-else sorguları yapılarak ailenin yardım alıp almayacağı hakkında bilgi sahibi olunması sağlanır. Şekil 5.10’da görülmektedir.

YARDIM SONUCU

• Yardım sonucu 'EVET'

Yerel Ağırlıklı Öğrenme

Yerine Koyarak Örnekleme

Sade Bayes Sınıflandırıcı

Şekil 5.10. Yardım Sonuç Ekranı

İşlemlerin tümü sağlandığında aile yardım alıyorsa “Yardım sonucu EVET” ya da aile yardım almıyorsa “Yardım sonucu HAYIR” olarak bir sonuç dönerek ailenin yardım alıp almaması konusunda bilgi sahibi olunur.

Yazılım genel olarak, web ortamından ulaşılabilecek şekilde oluşturulmakta ve bir web sunucusu üzerinden sonuçlar gönderilerek bir html sayfasında görüntülenmektedir. Sisteme giriş sağlayan kullanıcılar geliştirilen yazılıma kullanıcı adı ve şifresini kullanarak herhangi bir internet tarayıcısından web ortamında erişim sağlamaktadır. Uygulama verileri, veri tabanında tutulmakta ve işlemler bu veri tabanı üzerindeki verilerden alınıp gerçekleştirilmektedir. Yazılım birden çok kullanıcının giriş yapabileceği çoklu kullanıcı ve yetkilendirme yapısıyla çalışmaktadır. Girişi sağlayabilen kullanıcılar için ek olarak bir yazılım veya donanım kurulumu gerekmemektedir.

Özet olarak, kullanıcı giriş ekranıyla giriş sağlanıp Yeşilyurt Belediyesi sosyal Yardım İşleri’ne başvuran ailenin yardım alıp almayacağına dair karar verilip komisyona yardımcı olunabilmesi için bilgilerinin bilirkşi tarafından girilebildiği aynı zamanda girilen bilgilerin kontrolünün sağlandığı ve Karar Destek Mekanizması ile yardım sonucunun görülebildiği bir yapı oluşturulmuştur.

5.5 Karar Destek Sisteminin Test ve Bakımı

Geliştirme kısmı tamamlandıktan sonra test aşamasına geçilmiştir. Geliştirme aşamasında kodlaması tamamlanan Karar Destek Sistemi yazılım testine tabi tutularak kodlama esnasında oluşan hatalar test edilmiş ve daha sonra Karar Destek Sisteminin istenilen şekilde yapılıp yapılmadığını görmek amacıyla yeterlilik testi uygulanmıştır.

Yazılım ve yeterlilik testlerinin tamamlanmasıyla sistemi t mden sınamak i in kara kutu testi uygulanmıřtır. İlkel ve  ok kullanılan test y ntemlerinden olan kara kutu testi yazılım kodlarından bağımsız olarak ger ekleřmekte ve girdi verilerine g re sistemin tepkisini g z  n ne almaktadır. Bařka bir ifadeyle kara kutu testi kodlama, tasarım vb. dikkate almadan sistemin girdi ve  ıktı deęerlerinin istenilen gereksinimi karřılayıp karřılamadıęını inceleyen bir sınama olarak tanımlanmaktadır.

Bakım ařaması ise yazılım daęıtılıp, kullanılmaya bařlanmasıyla ger ekleřen bir ařamadır. Bu ařamanın amacı yazılımın iřlerlięinin devamını saęlayabilmektir ve bu ama la kolay kod hatalarının d zeltilmesi vb. yazılım deęiřiklikleri yapılarak yazılım bakımı ger ekleřmektedir. Karar Destek Sistemi, kullanıma hazır hale getirildikten sonra kullanılmaya bařlamadıęından dolayı bakım ařamasına ge ememiřtir.

6. SONUÇLAR VE DEĞERLENDİRME

Türkiye’de veri madenciliği üzerindeki ilgi giderek artmakta ve kullanım alanları her geçen yıl genişleyerek yayılmaktadır. Veri madenciliği eğitim, sağlık, bankacılık, tıp ve sosyal yardım gibi birçok alanda kullanılmaktadır. Veri madenciliğinin ülkemizde birçok alanlarda kullanılmasıyla beraber çeşitli kurum ve kuruluşlara fayda sağlamaktadır.

Ülkemizde meydana gelen işsizlik, fakirlik, muhtaçlık vb. sorunlar belediyelerin sosyal yardım ve sosyal hizmet alanında vazife üstlenmesine sebep olmuştur. Bu nedenle belediyelerin, sosyal yardım dağıtımı esnasında eldeki verileri hızlı bir şekilde kullanabilmesi ve ilgili komitede doğru karar verilmesi açısından Malatya Büyükşehir ilçe belediyesi olan Yeşilyurt Belediyesi örneği üzerinden alınan verilere veri madenciliği yöntemleri uygulanmış ve bir Karar Destek Mekanizması oluşturulması amaçlanmıştır.

Yapılan uygulama sonucunda, verileri en iyi başarımla sınıflandıran sınıflandırma algoritmaları karşılaştırılmış ve %81.8 başarımla Yerel Ağırlıklı Öğrenme algoritması en iyi sonucu veren algoritma olmuştur. Bu sonuç, belediyelerin sosyal yardım hizmetleri dağıtımında karar verme amaçlı olarak veri madenciliği alt yapısı olan bir karar destek sistemi kullanılabileceğini ortaya çıkarmıştır.

Tez kapsamında gerçekleştirilen Karar Destek Sistemi ile Yeşilyurt Belediyesi’nin kullandığı anket formuna uygun olarak tasarlanan sorular ve cevapları, kullanışlı bir arayüz ile kayıt altına alınmaktadır. Bu veriler, sınıflandırma algoritmalarından en iyi sonucu veren Yerine Koyarak Örnekleme, Yerel Ağırlıklı Öğrenme ve Sade Bayes Sınıflandırma algoritmalarının karar yapılarına göre değerlendirilmektedir. Belirtilen işlem bu karar yapılarına göre çalışmakta ve yardım alma sonucunun ekranda görüntülenmesi sağlanmaktadır Böylece anket verisi bulunan ailenin yardım alıp almaması hakkında ilgili komisyon ön bilgi sahibi olmaktadır. Geliştirilen yazılım ile kullanıcılar, kolaylıkla verileri otomatik olarak işleyebilecek ve yardım sonucunu komisyona sunarak değerlendirme aşamasında büyük kolaylıklar elde edebilecektir.

7. KAYNAKLAR

- [1] **Zaiane, O.R.**, 1999. Principles of Knowledge Discovery in Database. <https://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/slides/ch0s.pdf>, 20 Mayıs 2016.
- [2] **Robbins, R.J.**, 1995. Database Fundamentals. <http://www.esp.org/db-fund.pdf>, 21 Mayıs 2016.
- [3] **Ozkan, Y.**, 2013. Veri Madenciliği. Papatya Yayıncılık Eğitim, İstanbul.
- [4] **Fakı, B.M.**, 2015. Veri Madenciliği Yöntemlerini Kullanarak Anemi Sınıflandırılmasına Yönelik Bir Uygulama, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [5] **Karabatak, M., & İnce, M. C.**, 2004. Apriori algoritması ile öğrenci başarıları analizi. Elektrik Elektronik Bilgisayar Mühendisliği Sempozyumu, Bursa.
- [6] **Koyuncugil, A. S., & Özgülbaş, N.**, 2009. Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. International Journal Of Informatics Technologies, 2(2).
- [7] **Emel, G. G., & Taşkın, Ç.**, 2005. Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması. Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 6(2).
- [8] **Mocan, G.**, 2016. Perakendecilikte veri madenciliği uygulamaları ve sorunları, *Yüksek Lisans Tezi*, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [9] **SADIQMAL, F.**, 2015. Implementation of Some Medical Data Using Apriori Algorithm, *Yüksek Lisans Tezi*, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya.
- [10] **Zengin, E., Şahin, A., & Özcan, S.**, 2012. Türkiye’de Sosyal Yardım Uygulamaları, Celal Bayar Üniversitesi, Manisa, Yönetim, 133.
- [11] **Şener, Ü.**, 2010. Yoksullukla Mücadelede Sosyal Güvenlik, Sosyal Yardım Mekanizmaları ve İş Gücü Politikaları. TEPAV Politika Notu.
- [12] **Pektaş, E. K.**, 2010. Türkiye’de Sosyal Belediyecilik Uygulamaları ve Temel Sorunlar. Akademik İncelemeler Dergisi, 5(1).
- [13] **Cao, L.**, 2012. Social security and social welfare data mining: An Overview. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 837-853.

- [14] **Kum, H. C. M., Duncan, D., Flair, K., & Wang, W.**, 2003. Social welfare program administration and evaluation and policy analysis using knowledge discovery and data mining (kdd) on administrative data. In *Proceedings of the 2003 annual national conference on Digital government research* (pp. 1-6). Digital Government Society of North America.
- [15] **Anderson, R., & Mansingh, G.**, 2014. Data Mining Approach to Decision Support in Social Welfare. *International Journal of Business Intelligence Research (IJBIR)*, 5(2), 39-61.
- [16] **Chan, A.**, Introduction to Computers for the Arts and Social Sciences. <http://people.scs.carleton.ca/~achan/>, 21 Mayıs 2016.
- [17] **Zaki, M.J. and Meira JR.W.**, 2014. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, New York.
- [18] **Rajaraman, A., Leskovec, J. and Ullman, J.D.**, 2014. Mining of Massive Datasets. Cambridge University Press, United Kingdom.
- [19] URL-1, <https://www.emaze.com/@AQLRLFLR/untitled>, Web Madenciliği: Web Sayfalarında Örüntü Keşfi. 24 Aralık 2017.
- [20] **Cohen, W.W. and Richman, J.**, 2002. Learning to Match and Cluster Large High-Dimensional Data Sets for Data Integration, *Edmonton, Alberta, Canada*, **1**, 58113-567.
- [21] **Hu, Q., Yu, D. and Xie, Z.**, 2005. Information-Preserving Hybrid Data Reduction Based on Fuzzy-Rough Techniques, *Pattern Recognition Letters*, 27, 414-423.
- [22] **Fayyad, U., Shapiro, G.P. and Smyth P.**, 1996. From Data Mining to Knowledge Discovery in Databases, *American Association for Artificial Intelligence*, 38, 0738-4602.
- [23] **Oliveira, S.R.M. and Zaiane, O.R.**, 2010. Privacy Preserving Clustering by Data Transformation, *Journal of Information and Data Management*, **1**, 37-51.
- [24] **Jiang, N. and Gruenwald, L.**, 2006. Research Issues in Data Stream Association Rule Mining, *SIGMOD Record*, **35**, 13-19.
- [25] URL-2, <http://web.itu.edu.tr/~sgunduz/courses/verimaden/slides/d4.pdf>, Veri Madenciliği Farklı Sınıflandırma Yöntemleri. 15 Eylül 2016
- [26] **Coşkun, C., & Baykal, A.** 2011. Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması. *Akademik Bilişim*, Malatya.
- [27] **Aşuroğlu, T.** 2015. Hitit çivi yazısı işaretlerinin bilgisayar desteği ile okunması ve veri madenciliği uygulama örnekleri, *Yüksek Lisans Tezi*, Başkent Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

- [28]URL-3,http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html , Confusion Matrix. 30 Eylül 2016
- [29]URL-4, https://en.wikipedia.org/wiki/Confusion_matrix, Confusion Matrix. 23 Aralık 2016
- [30] **Xu, M., Watanachaturaporn, P., Varshney, P.K. and Arora, M.K.**, 2005. Decision Tree Regression for Soft Classification of Remote Sensing Data, *Remote Sensing of Environment*, **97**, 322 – 336.
- [31] **Anyanwu, M.N. and Shiva, S.**, 2009. Comparative Analysis of Serial Decision Tree Classification Algorithms, *International Journal of Computer Science and Security*, **3**, 230-240.
- [32] **Han, J., Kamber, M. and Pei, J.**, 2012. Data Mining: Concepts and Techniques, Morgan Kaufmann Publications, USA.
- [33] **Baradwaj, B.K. and Pal, S.**, 2011. Mining Educational Data to Analyze Students' Performance, *International Journal of Advanced Computer Science and Applications*, **6**, 63-68.
- [34] **Sheraz, M., Dedu, S. and Preda, V.**, 2015. Entropy Measures for Assessing Volatile Markets, *Procedia Economics and Finance*, **22**, 655-662.
- [35] **Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R. and Honrao, V.**, 2013. Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms, *IJDKP*, **5**, 39-52.
- [36] **Nalçakan, Y., Bayramoğlu, Ş. S., & Tuna, S.**, 2015. Sosyal Medya Verileri Üzerinde Yapay Öğrenme ile Duygu Analizi Çalışması.
- [37]URL-5,https://en.wikipedia.org/wiki/C4.5_algorithm C4.5 Algoritması.6 Haziran 2016.
- [38] **Özhan, E.**, 2014. Yapay Zeka Yöntemleri İle Dünya Depremlerinin Modellenebilirliği, *2nd International Symposium On Innovative Technologies In Engineering And Science*, Karabük Üniversitesi, 18-20 Haziran.
- [39]URL-6,<http://bilgisayarkavramlari.sadievrenseker.com/2012/11/13/c4-5-agaci-c4-5-tree/>, C4.5 Ağacı. 22 Haziran 2016.
- [40] **Diplars, S., Tsoumakas, G., Mitkas, P.A. and Vlahavas I.**, 2005. Protein Classification with Multiple Algorithms, *Springer-LNCS* **3746**, pp. 448-456.
- [41]URL-7, <http://ieeexplore.ieee.org/abstract/document/844072/>, Real-time robot learning with locally weighted statistical learning. 20.10.2017
- [42] **Triola, M.F.**, 2010. Bayes' Theorem in *Elementary Statistics*, 11. <http://faculty.ccp.edu/faculty/sgramlich/152/bayes.pdf>, 30 Mayıs 2016.

- [43]URL-8,<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/learn-43/lib/photoz/.g/web/lwr.html>, Locally weighted Learning. 20.10.2017
- [44]URL-9,<http://aass.oru.se/~tdt/ml/extra-slides/LWLforcontrol.pdf> Locally Weighted Learning for Control. 20.10.2017
- [45] Englert, P., 2012. Locally Weighted Learning. *In Seminar Class on Autonomous Systems*, Department of Computer Science TU Darmstadt.
- [46] Christopher, A., Andrew, M., & Stefan, S., 1997. Locally weighted learning. *Artif Intell Rev*, 11(1-5), 11-73.
- [47] **Sayin, K. E.**, 2013. Importance of Data Preprocessing For Improving Classification Performance on CAC Data Set, *Yüksek Lisans Tezi*, Kadir Has Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul
- [48] **Cihan, P.**, 2013. Veri madenciliği ile yazılım mühendisliği dersi projelerinin iyileştirilmesi, *Yüksek Lisans Tezi*, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [49] **Platt, J.**, 1998. Sequential minimal optimization: A fast algorithm for training support vector machines.
- [50] **Karhan, Z.**, & Karhan, M., 2016. Görüş madenciliğinde makine öğrenmesi algoritmalarının performans değerlendirmesi. *Türk Doğa Ve Fen Dergisi*, 51-54.
- [51] **BİLGİN, M.**, 2017. Gerçek Veri Setlerinde Klasik Makine Öğrenmesi Yöntemlerinin Performans Analizi. *Breast*, 2(9), 683. *Akademik Bilişim*, Bursa.
- [52] **Kılınç, D., Borandağ, E., Yücalar, F., Özçift, A., & Bozyiğit, F.**, 2015. Yazılım Hata Kestiriminde Kolektif Sınıflandırma Modellerinin Etkisi, *IX Ulusal Yazılım Mühendisliği Sempozyumu*, Yaşar Üniversitesi, Bornova-İzmir, 9-11 Eylül.
- [53]URL-10, https://en.wikipedia.org/wiki/Bootstrap_aggregating, Bagging Algoritması. 31 Mayıs 2016.
- [54] **Amasyalı, M. F., & Ersoy, O.**, 2011. Sınıflandırıcı Topluluklarında Başarı Tabanlı Budama ve Ağırlıklı Oylama. *audiology*, 70(5), 169.
- [55] **Amasyalı, M. F.**, 2014. Örneklemeli K-ortalama Algoritması, Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü.
- [56] **Yılmaz, H.**, 2014. Random Forests yönteminde kayıp veri probleminin incelenmesi ve sağlık alanında bir uygulama, *Yüksek Lisans Tezi*, Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.
- [57]URL-11,<http://home.etf.rs/~vm/os/dmsw/Random%20Forest.pptx>, Random Forest Algoritması. 22 Haziran 2016.

- [58]URL-12,[http://www.academia.edu/25049513/Random Forest Rastgele Orman Algoritmas %C4%B1](http://www.academia.edu/25049513/Random_Forest_Rastgele_Orman_Algoritmas_%C4%B1), Random Forest Algoritması. 22 Haziran 2016.
- [59]URL-13,<https://tr.wikipedia.org/wiki/Weka>, Weka. 2 Ocak 2017.
- [60] **Daş, B., & Varol, A.**, 2013. 2D: 4D Sayısal Parmak Oranına Göre Bireylerin Kişilik Durumlarının Sınıflandırılması, 1 st International Symposium on Digital Forensics and Security (ISDFS'13), Elazığ,20-21 Mayıs.

