



**TÜRKİYE'DE ANLAŞMALI BOŞANMA İLE SOSYAL GÜVENLİK
KURUMUNDAN AYLIK ALANLAR ÜZERİNE VERİ MADENCİLİĞİ
CART ALGORİTMASI UYGULAMASI**

Nihal YAŞASINOĞLU

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

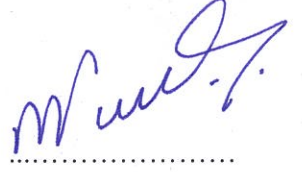
TEMMUZ 2019

Nihal YAŞASINOĞLU tarafından hazırlanan “TÜRKİYE’DE ANLAŞMALI BOŞANMA İLE SOSYAL GÜVENLİK KURUMUNDAN AYLIK ALANLAR ÜZERİNE VERİ MADENCİLİĞİ CART ALGORİTMASI UYGULAMASI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İstatistik Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Meral EBEGİL

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

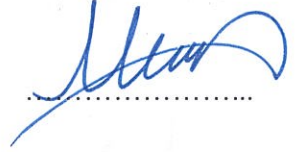
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.



Başkan: Prof. Dr. Mehtap AKÇİL OK

Beslenme ve Diyetetik Programı, Başkent Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.



Üye: Prof. Dr. Fikri GÖKPINAR

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.



Tez Savunma Tarihi: 23/07/2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ

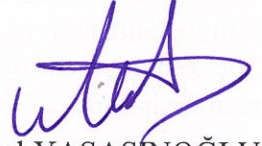
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.



Nihal YAŞASINOĞLU

23/07/2019

TÜRKİYE'DE ANLAŞMALI BOŞANMA İLE SOSYAL GÜVENLİK KURUMUNDAN AYLIK ALANLAR ÜZERİNE VERİ MADENCİLİĞİ CART ALGORİTMASI

UYGULAMASI

(Yüksek Lisans Tezi)

Nihal YAŞASINOĞLU

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Temmuz 2019

ÖZET

Günümüzde bilgisayar dosyalarında ve veri tabanlarında tutulan verilerin miktarı çarpıcı bir oranda artmaktadır. Aynı zamanda, veri tabanında tutulan bu verileri kullananlar onlardan daha karmaşık bilgiler beklemektedir. Basit sorgulama teknikleri, artan bu bilgi talebini desteklemekte yetersiz kalmakta, bu ihtiyaçları gidermek için veri madenciliği teknikleri kullanılmaktadır. Veri madenciliği farklı araştırmaları başarmak için pek çok farklı algoritma içermektedir. Tüm bu algoritmalar veriye bir model uydurmak amacı taşımaktadır. Veri madenciliğinde esas, önceden bilinmeyen, akla gelmeyen, tahmin edilemeyen ortaya koymaktır. Anlaşmalı boşanmalar sorunu Türkiye sosyal güvenlik mevzuatında yer alan boşluğun, sigortalılarca suistimal edilmesi sonucu ortaya çıkmaktadır. Mevzuatta yer alan mevcut hükümlere göre, yetim kalan çocukları sosyal güvence çatısı altına almayı hedeflemesine rağmen, evlendikten sonra boşanan ileri yaştaki kız çocukları da yetim olarak nitelendirildiklerinden, babadan veya anneden dolayı ölüm aylığına hak kazanmaktadır. Türkiye’de evlendiği halde ölen ana veya babasının aylığını almak isteyen yetim kız çocukları evli oldukları eşleriyle anlaşmalı olarak boşanarak aynı evde yaşamaktadırlar. 5510 sayılı kanunun gelir ve aylık bağlanmayacak haller başlıklı 56 ncı maddesinin son fıkrasında eşinden boşandığı halde, boşandığı eşiyle fiilen birlikte yaşadığı belirlenen eş ve çocukların, bağlanmış olan gelir ve aylıkların kesileceği hüküm altına alınmıştır. Buradan hareketle, anlaşmalı boşandığı tespit edilen kişilerin davranış modeli ortaya konularak, bugüne kadar tespit edilemeyen grubun özelliklerini belirleyerek, denetim birimlerinin karar vermesinde yardımcı olacak nitelikte sonuçlar ortaya koymak araştırmanın temel amacını oluşturmaktadır.

Bilim Kodu : 20514

Anahtar Kelimeler : Veri madenciliği, anlaşmalı boşanma, ölüm aylığı, CART

Sayfa Adedi : 57

Danışman : Doç. Dr. Meral EBEGİL

DATA MINING CART ALGORITHM IMPLEMENTATION ON PENSIONERS FROM
SOCIAL SECURITY INSTITUTION WITH FAKE DIVORCE IN TURKEY

(M. Sc. Thesis)

Nihal YAŞASINOĞLU

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2019

ABSTRACT

Today, the amount of data kept in computer files and databases is increasing dramatically. At the same time, those who use this data stored in the database are waiting for more complex information. Simple inquiry techniques fail to support this increasing demand for information, and data mining techniques are used to address these needs. Data mining contains many different algorithms to accomplish different researches. All these algorithms are intended to fit a model into the data. In data mining, it should be to reveal the unpredictable. Fake divorce problem of space located in Turkey's social security legislation, are emerging as a result of the abuse of Insurers. According to the current provisions of the legislation, the female children who are divorced after marrying are considered to be orphans and they are entitled to orphans pension due to father or mother. As they can entitle to orphans pension due to father or mother, after orphan girls get fake divorce live in the same house with their ex husband. Last paragraph of article 56 of 5510 numbered law point out that if orphan girls determined to fake divorced and live together with their ex spouse, the income and pensions will be cut. From this point of view, the main aim of this study is to determine the behavioral model of the people who are found to be fake divorce and to determine the characteristics of the group which has not been determined until now. To provide that will help the decision making of the audit units.

Science Code : 20514

Key Words : data mining, fake divorce, survivors pension, CART

Page Number : 57

Supervisor : Assoc. Prof. Dr. Meral EBEGİL

TEŞEKKÜR

Yüksek lisans çalışmalarım süresince yardım ve katkılarını benden esirgemeyen değerli hocam tez danışmanım Doç. Dr. Meral EBEGİL'e, teşekkür ederim. Çalışmalarımı yürütmemde bana her daim destek veren Sosyal Güvenlik Uzmanı eşim Hamza YAŞASINOĞLU'na, bu süreçte aramıza katılan sevgili oğlum Mert'e ve bu güne kadar hep yanımda olan sevgili aileme manevi desteklerinden dolayı ayrıca teşekkür ederim.



İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
RESİMLERİN LİSTESİ	x
HARİTALARIN LİSTESİ.....	xi
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ.....	1
2. VERİ MADENCİLİĞİ.....	5
2.1. Veri Madenciliği Alanında Yapılan Çalışmalar.....	6
3. VERİ MADENCİLİĞİNDE KULLANILAN ALGORİTMALAR	11
3.1. Sınıflama Algoritmaları	11
3.1.1. ID3 algoritması	12
3.1.2. C4.5 algoritması.....	14
3.1.3. CART algoritması	15
3.1.4. SPRINT algoritması.....	16
3.1.5. Bayesyen sınıflama	17
3.1.6. K-En yakın komşu.....	18
3.1.7. Yapay sinir ağları	19
3.2. Kümeleme Algoritmaları	20
3.2.1. En yakın komşu algoritması (SLINK).....	20
3.2.2. En uzak komşu algoritması (Cure).....	21
3.2.3. CHAMELEON algoritması.....	21

	Sayfa
3.2.4. K-Ortalama algoritması	22
3.2.5. PAM algoritması	23
3.3. Birliktelik Kuralları	24
3.3.1. Apriori algoritması	24
3.3.2. AIS algoritması	25
4. TÜRKİYEDE ANLAŞMALI BOŞANMALAR.....	27
4.1. Ölüm Aylığına Hak Kazanma Şartları	27
4.2. Anlaşmalı Boşanma	28
4.3. Anlaşmalı Boşanmanın Tespiti	29
5. UYGULAMA.....	31
5.1. Uygulamada Kullanılan Değişkenlerin Tanımları	31
5.2. Uygulamada İzlenen Adımlar	38
5.3. Karar Ağacının Elde Edilmesi	39
6. SONUÇ VE ÖNERİLER	49
KAYNAKLAR	53
ÖZGEÇMİŞ	57

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 5.1. İllerin bulunduğu bölgelere göre belirlenen kategoriler	32
Çizelge 5.2. Yaş aralıklarına göre kategoriler	33
Çizelge 5.3. Aylık türü değişkenine verilen kodlar	34
Çizelge 5.4. Ölüm-boşanma süresinin kategorik veriye dönüştürülmesi	35
Çizelge 5.5. Prim ödeme gün sayısının kategorik veriye dönüştürülmesi	36
Çizelge 5.7. Sektör sınıflaması	37
Çizelge 5.8. Sınıflama sonucunda elde edilen karar ağacında oluşan davranış profilleri	41

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 3.1. Yapay sinir ağı.....	19



HARİTALARIN LİSTESİ

Harita	Sayfa
Harita 5.1. Türkiye'nin 7 bölgesinin gösteren harita	32



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar

Açıklamalar

CART

Classification and Regression Tree

CHAID

Chi-Squared Automatic Interaction Detector

HBA

Hane Halkı Bütçe Araştırması

KNN

K-Nearest Neighbor

PAM

Partitioning Around Medoids

TÜİK

Türkiye İstatistik Kurumu

YMA

Yaşam Memnuniyet Anketi

1. GİRİŞ

Bilgisayar dosyalarında ve veri tabanlarında tutulan verilerin miktarı çarpıcı bir oranda artmaktadır. Aynı zamanda, veri tabanında tutulan bu verileri kullananlar onlardan daha karmaşık bilgiler beklemektedir. Örneğin ticaret yapan biri artık basit ticari listelerle tatmin olmamaktadır. Bunun yerine tüketicilerin geçmişte satın aldığı ürünlerden yola çıkarak gelecekte satın alacakları ürünlerin tahmini hakkında daha detaylı bilgi istemektedir. Basit sorgulama teknikleri, artan bu bilgi talebini desteklemekte yetersiz kalmaktadır. Bu ihtiyaçları karşılamak amacıyla veri madenciliği teknikleri yaygın olarak kullanılmaktadır.

Yaşamımızın her anında yaptığımız işlemler kayıt altına alınmakta ve yığın oluşturmaktadır. Örneğin internette ya da marketlerde yapılan alışverişler, laboratuvarlarda sonuçlanan tahliller, kamu kurumlarında yapılan işlemler, bankalarca yapılan işlemler gibi. Çok büyük yığınlara hizmet veren bu kurumların veri tabanlarında yığına ait özellikler ve yapılan işlemler kayıt altında tutulmaktadır. Bu yığınlarda ortaya çıkarılmayı ve değerlendirilmeyi bekleyen, maden değerinde veriler bulunmaktadır. Burada da veri madenciliğine ihtiyaç duyulmaktadır.

Veri madenciliği günümüzde daha çok işletmelerin ürünlerini müşterilerine daha etkin şekilde pazarlayabilmeleri için müşterilerin davranış modellerini çıkarmak amacıyla kullanılmaktadır. Ngai, Xiu ve Chau (2009) çalışmalarında veri madenciliğinin müşteri ilişkileri yönetimi (customer relationship management (CRM)) konusunda literatürde eksik olduğu düşünülerek 2000-2006 yılları arasında 24 dergide yayımlanan 87 makale seçilmiş ve bu makaleler müşteri ilişkileri yönetimi açısından 4 başlıkta (müşteri tanımlama, müşteri ilgisi, müşteri tutma ve müşteri artırma), uygulanan veri madenciliği teknikleri bakımından da 7 başlıkta (association, classification, clustering, forecasting, regression, sequence discovery, visualization) kategorize edilerek bu konudaki ilk akademik çalışma olan literatür incelemesi yapılmıştır. Yapılan çalışma sonucunda araştırmacıların en çok müşteri tutma konusuna ilgi gösterdikleri görülmüştür. Diğer yandan müşteri ilişkileri yönetimi konusunda sınıflama (classification) ve ilişki kuralları (association rule) algoritmalarının en çok kullanılan modeller olduğu görülmüştür. Shafer, Agrawal ve Mehta (1996) veri madenciliği için sınıflamanın önemli bir problem olduğunu ifade ederek SPRINT adında yeni bir algoritma geliştirmişlerdir. Sınıflama problemi

açısından iyi çalışan bugün kullanılan sınıflama algoritmaları, veri setindeki tüm ya da bir kısım verilerin bellekte kalmasını gerektirmektedir. Bu durum büyük veri tabanlarında veri madenciliğinin uygulanabilirliğini kısıtlamaktadır. Bunun sonucunda ölçeklenebilir, hızlı ve tüm bellek kısıtlamalarını ortadan kaldıran karar ağacı sınıflamasına dayanan SPRI^{NT} adında yeni bir karar ağacı algoritması önerilmiştir. Du ve Zhan (2003) veri madenciliğinde önemli bir hususun da kişisel verilerin mahremiyeti konusu olduğunu ifade ederek bu konuda rasgele tepki teknikleri(Randomized Response Techniques) adı altında gizlenmiş veriden karar ağacı sınıflayıcıları için bir yöntem geliştirilmişlerdir. Bu metodun doğruluğu, maskelenmemiş veriden elde edilen karar ağaçlarıyla deney yapılarak test edilmiş, sonuç olarak veri maskelense dahi önerilen metodun doğruluğunun oldukça yüksek olduğu gösterilmiştir. Chen ve diğerleri (1996) veri madenciliği tekniklerini veri tabanı araştırmacılarının gözünden inceleyerek sınıflandırmış ve bu tekniklerle ilgili karşılaştırmalı bir çalışma sunmuştur.

Bu çalışmada ise veri madenciliği sınıflama tekniklerinden uygulamalarda en çok tercih edilen metotlardan CART algoritması ile Sosyal Güvenlik Kurumundan temin edilen gerçek bir veri seti üzerine sınıflama yapılmıştır. Anlaşmalı olarak eşinden boşanarak anneden veya babadan aylık alırken tespit edilen yetim kız çocukları veri setini oluşturmuştur. Bu veri setindeki özelliklere göre sınıflama işlemi yapılmıştır. Araştırmada, denetim personellerinin denetim süreçlerine yardımcı olunması amacıyla anlaşmalı boşananların ortak özelliklerini ortaya çıkarmak amaçlanmıştır.

Çalışmanın ikinci bölümünde, veri madenciliğinin tanımı ve uygulandığı alanlar hakkında bilgiler verildikten sonra yapılan çalışmalardan bahsedilmiştir. Burada veri madenciliğinin çok çeşitli alanlarda kullanılabildiği görülebilir.

Araştırmanın üçüncü bölümünde, veri madenciliği tekniklerinin hangi tanımlar altında toplandığı (Sınıflama, Kümeleme, Birliktelik Kuralları) kısaca açıklanarak bu başlıklar altında bulunan algoritmalarından bazıları, hangi amaçlarda kullanıldığı, teorisi, avantajları, dezavantajları ve bazıları için de örnekler verilerek açıklanmaya çalışılmıştır.

Dördüncü bölümde, sosyal güvenlik mevzuatında ölüm aylıklarına ne şekilde hak kazanılacağı ve Türkiye’de anlaşmalı boşanmalar anlatıldıktan sonra bu kişilerin tespitine

yönelik mevcut durumda yapılan bürokratik işlemler ile denetim sürecinin ne kadar zahmetli ve zaman alıcı olduğu açıklanmıştır.

Beşinci bölümde ise Türkiye’de anlaşmalı boşanarak anne veya babasından yetim aylığı alırken tespit edilen, gelir veya aylıkları kesilen kişilere ait kullanılan değişkenler tanımlanmıştır. Daha sonra bu değişkenlerle veri madenciliği sınıflama algoritmalarından en çok tercih edilen yöntemlerden olan CART algoritması (Classification and Regression Tree-Sınıflama ve Regresyon Ağacı) kullanılarak karar ağacı oluşturulmuş ve bu kapsamdaki kişiler ortak özelliklerine göre sınıflandırılmıştır. Karar ağacından elde edilen sonuçlar yorumlandıktan sonra elde edilen bulgular değerlendirilmiştir.



2. VERİ MADENCİLİĞİ

Veri madenciliği; genellikle veri tabanında saklı olan gizli bilgiyi ortaya çıkarmak olarak tanımlanmaktadır. Alternatif şekilde tündengelimsel veri analizi anlamına gelmektedir. Veri madenciliği farklı araştırmaları kullanmak için pek çok farklı algoritma içermektedir. Tüm bu algoritmalar veriye bir model uydurmak amacı taşımaktadır. Bu algoritmalar verileri incelemekte ve incelenen verilerin karakteristiğine en uygun modele karar vermektedir. Veri madenciliğinde esas önceden bilinmeyi, akla gelmeyi, tahmin edilemeyi ortaya koymak olmalıdır. Veri madenciliğinin şuana kadarki gelişimi; veri tabanları, istatistikler ve makine öğrenmeyi içeren (machine learning) pek çok disiplinden yıllar içinde etkilenecek ortaya çıkmıştır. Veri madenciliği, verinin bütünü kullanması bakımından diğer istatistiksel analizlerden ayrılmaktadır. Bu yöntemlerle, geleneksel yollarla elde edilmiş küçük verilerle çalışmak yerine daha kolay değerlendirme yapılabilir. Veriler arasındaki anlamlı yeni korelasyonları ve örüntüleri keşfetme süreci de veri madenciliği olarak tanımlanır. Veri madenciliği, geniş gözlemleri analiz etme, yeni yollarla veri sahiplerine anlamlı ve kullanışlı bir şekilde özetleme yöntemi diye de tanımlanabilir (Dunham, 2003: 2, 5).

Veri madenciliğinin uygulama alanlarını sıralayacak olursak; Bunların başında pazarlama yöntemi gelir. Müşterilerin satın alma stratejilerinin belirlenmesi, yeni müşteri kazanımı, pazar sepeti analizi, makarna alan müşterinin sos-ketçap alması örneği, risk yönetimi ve dolandırıcılık saptamadır. Örneğin; Araştırmada Farmer Group şirketinin “Risk Analiz Yönetim” paketi; sigorta oranı belirlenmesi, yatırım portföy yönetimi, iyi ve kötü kredi riskleri taşıyan şirketlerin belirlenmesinde, Twentieth Century Fox adlı film şirketi ise fatura bilgilerinin analizinde hangi aktör hangi film hangi ilde en çok izleniyor? Sorusuna yanıt bularak yeni film projeleri için strateji belirlemede veri madenciliğini kullanmıştır.

Veri madenciliği, pazarlama dışında, tıp alanında da birçok hastalığın teşhisinde kullanılmaktadır. Örneğin gebelikte down sendromu riski anne karnından alınacak sıvının incelenmesi ile olasılığı yüksek bir tanı ortaya koymaktadır. Biyoloji alanında DNA analizi de veri madenciliğinin kullanımına bir örnektir.

2.1. Veri Madenciliği Alanında Yapılan Çalışmalar

Zhang, Zhang ve Yang (2003) yaptıkları çalışmada veri hazırlamanın, veri madenciliğinin uygulanmasında önemli bir aşama olduğunu, çeşitli veri kaynaklarından ve internetten pek çok bilgiye erişilebildiğini ancak bu bilgilerin çoğu düşük kaliteli olduğunu ifade etmişlerdir. Pek çok şirket ve organizasyon, veri madenciliği teknikleriyle yüksek kalitede bilgi sağlayabilmek için bu verileri nasıl dönüştürmesi gerektiğiyle ilgilendiği vurgulanarak veri hazırlama teknolojileri altı kategoride ifade edilmiştir. Bunlar; veri tabanından bilgi çıkarma sürecindeki teknikleri bir araya getiren hibrid madencilik sistemleri (hybrid mining systems), istatistiksel tekniklerle verileri kümelemek (data clustering), çok büyük bir bilgi kaynağı olan internetten bilgiyi toplayan dağıtan paylaşan bir yapı olan yapay zeka web (Web intelligence), verileri özelliklerine göre seçen metodlar (C4.5 algoritması), kayıp değerleri doldurmak ve bir konuda çok az bilgi olduğunda o konu ile bağlantılı dış veri tabanlarından sağlanacak bilgi şeklinde tanımlanmıştır.

Baradwaj ve Pal (2011) öğrencilerine daha kaliteli eğitim sunmak isteyen eğitim kurumları için, eğitim verilerinde saklı olan bilgileri ortaya çıkarmak amacıyla veri sınıflama teknikleri ve karar ağacı yöntemlerinden ID3 algoritması kullanarak öğrencilerin başarı durumlarındaki düşüşü tanımlamış ve ayrıca özel ilgi gösterilmesi gereken öğrencileri tanımlayarak uygun rehberlik hizmetinin bu öğrencilere ulaştırılmasını sağlamıştır.

Ülkemizde veri madenciliği konularında yapılan araştırmalardan bazıları ise şöyledir: BUDAK (2012) Bilgisayar okuryazarlığını etkileyen faktörlerin belirlenmesi amacıyla Kocaeli Üniversitesi'nde 2012-2013 öğretim yılı güz döneminde Temel Bilgi Teknolojileri Kullanımı (TBTk) dersini alan öğrenciler arasından rastgele erişim yöntemiyle seçilen örnek yığına bir anket düzenlemiş ve sonuçları veri madenciliğinden yararlanarak, program yardımıyla analiz etmiştir. Çalışmadan elde edilen sonuçlara göre, ailelerin gelir düzeyinin bilgisayar deneyim süresiyle ilişkisinin yüksek, ancak bilgisayar sahipliği ile ilişkisinin düşük olduğu görülmüştür. Analiz sonucunda öğrencilerin bilgisayar okuryazarlığı konusunda ön yargılarının son derece hatalı olduğu, ancak ders içerisinde gösterilen müfredatın, bilgisayar okuryazarlığını olumlu yönde etkilediği görülmüştür. Araştırma sonucunda, bilgisayar okuryazarlığı ile ilgili çalışmaların erken dönem eğitim süreçlerine katılmasının gerekliliği ve öğrencilerin bilinçli kullanıcılar haline gelmeleri için eğitim almalarının bilgisayar okuryazarlığında oldukça olumlu etkilerinin olacağı

söylenmiştir. Alan (2014) Cumhuriyet Üniversitesi İİBF öğrencilerine ait verileri kullanarak yaptığı karar ağaçları yöntemiyle veri madenciliği uygulanan araştırmasında, öğrencilere ait verilerden yararlanarak, hem bu verileri en başarılı sınıflandıran karar ağaçlarına ait algoritmayı, hem de bu algoritmanın üreteceği sınıflar tespit edilmeye çalışmıştır. Araştırmanın sonucunda LADTree algoritmasının öğrenci verilerini sınıflamada en başarılı algoritma olduğu ve 19 değişik sınıf ürettiği sonucuna varılmıştır. Bu analizler sonucunda, öğrencilere ait verilerden yararlanarak hangi sınıfa düşecekleri tahmin edilebilmiştir.

Emel ve Taşkın (2014) tarafından satış hacmi analizi için veri madenciliği karar ağacı yöntemi uygulanan bir diğer çalışmada işletmenin belirli bir dönemde yaptığı satışlar, müşterinin satın alma davranış özelliklerine dayalı olarak analiz edilmiştir. CART tekniği kullanılarak harcama tutarına göre müşteriler sınıflara ayrılarak yapılan analiz sonucunda müşterilerin tanımlanması, hedef müşteri grubunun belirlenmesi ve müşterilerin işletmeye bağlılığının güçlendirilmesine yönelik etkin pazarlama stratejilerinin oluşumuna önemli katkıda bulunduğu gösterilmiştir. Ayrıca, pazarlama stratejileri envanter, dağıtım, müşteriye hızlı cevap verebilme gibi işletmenin bir çok politikasını olumlu yönde etkilediği tespit edilmiştir. Böylece işletmede maliyetlerin düşürülmesi sureti ile kar artışı meydana gelebildiği gösterilmiştir. Ayık ve diğerleri (2013) çalışmalarında, öğrencilerin mezun oldukları lise türleri ve lise mezuniyet dereceleri ile kazandıkları fakülteler arasındaki ilişkiyi veri madenciliği teknikleri kullanılarak incelemişlerdir. MineSet 3.2 programı ile Atatürk Üniversitesi'ni kazanan öğrencilerin büyük bir kısmının genel liselerden geldiği sonucuna varmışlardır ve lise türünün arzu edilen bir fakültenin kazanılmasında büyük öneme sahip olduğunu tespit etmişlerdir.

Gemici (2012) yaptığı çalışmada IMKB Ulusal pazarda işlem gören 10 şirketin hisse senedi endekslerinin artışları ve azalışları arasındaki birlikteliklerin ortaya çıkarılmasını amaçlamıştır. Çalışmada teknik olarak, birliktelik kuralı algoritmalarından Apriori algoritması kullanılarak, Spss Clementine 12.0 programı ile veriler arasında güçlü korelasyon olup olmadığı kontrol edilmiş ve yatırımcının kazancını artırıcı birliktelik kuralları ortaya konulmuştur. Aydın (2007) yaptığı çalışmada uzaktan eğitim sisteminin planlama faaliyetlerine katkı sağlayabilecek öğrenci performansını tahmin etmeye yönelik model geliştirmiş ve mezun olan öğrencilerin profillerini belirlemeye yönelik kümeleme çalışması yapmıştır. Öğrenci başarısını tahmin etmeye yönelik çalışmada C5.0 karar ağacı

algoritmasının kullanıldığı bir tahmin modeli “K-means” algoritması kullanarak beş küme elde edilmiş ve sonuçta elde edilen bilgilerin bilgisayar kullanımı ve öğrenci başarısı arasındaki ilişkiyi doğrular nitelikte olduğu görülmüştür. Çığışar (2017) Türkiye İstatistik Kurumu’ndan (TÜİK) alınan veriler ile WEKA yazılımı kullanarak yaptığı araştırmada, bankacılık ve sigortacılıkta büyük önem taşıyan kredi risklerini belirlemek ve temerrüde düşme olasılığını tahmin etmekte kullanılabilecek en iyi sınıflama algoritmasını belirlemiştir. Akça (2014) Gazi Üniversitesi öğrencilerinin profilini tespit ettiği araştırmasında veri madenciliği algoritmalarından CHAID, QUEST, CART ve C5.0 uygulamıştır. Analiz sonucunda not ortalamasının cinsiyet değişkeni üzerinde etkili olduğu ortaya konulmuştur. Demirel (2010) yapmış olduğu çalışmada karar ağaçları yönteminde kullanılan CHAID algoritması ile ülkemizde Sosyal Güvenlik Kurumu ile ilişkilendirmiştir. İşyerlerinde yapılandırma kanunundan yararlanan işyeri profillerini belirlemeye çalışmış ve daha sonra çıkarılacak olan kanunların bu yapılan çalışmada öngörülen bilgiler ışığında karar destek amaçlı olarak katkı sağlayacağını ifade etmiştir.

Diler (2016) yılında yaptığı yüksek lisans tez çalışmasında Türkiye İstatistik Kurumu tarafından gerçekleştirilen Yaşam Memnuniyet Anketi (YMA) sonuçlarına ayrı ayrı CART ve CHAID algoritmaları uygulayarak karşılaştırmıştır. Araştırma sonucunda ise CHAID algoritmasının CART algoritmasına göre daha araştırmacı bir yapıya sahip olduğunu göstermiştir. Veri madenciliği tekniklerini süreçleriyle beraber inceleyen ve gerçek bir veri setine uygulama yapan, Kocabaş (2010) internet servis sağlayıcısı firmasından sağladığı 2009 yılına ait ADSL hizmetini iptal eden abonelerin iptal nedenleri ve bunu etkileyen değişkenleri CHAID, C.5 ve Lojistik Regresyon algoritmaları ile incelemiş ve iptal nedenlerinin bireysel tercih ve ekonomik nedenler üzerine baskın olduğunu gözlemlemiştir. Çakır (2012) çalışmasında Küçük ve Orta Ölçekli İşletmeleri Geliştirme ve Destekleme İdaresi Başkanlığı Veri Tabanı’na kayıtlı işletmelerin gelişmelerini ve büyümelerini sağlayacak sistem ve süreçlerin belirlenmesi, işletmelerin uyguladıkları sistem ve süreçlerin birlikte görülme sıklıklarının tespit edilmesi ile hangi sistem ve süreçlerin işletmeleri başarıya taşıdığını belirlemek amacıyla, bu işletmelere ait veriler üzerine veri madenciliği, sınıflama ve birliktelik kuralları tekniklerinden CART ve yapay sinir ağları yöntemlerini kullanarak analiz etmiştir.

Kıran (2010) yapmış olduğu çalışmada penisilin türündeki antibiyotikleri kullananların karakteristik özellikleri Sosyal Güvenlik Kurumu ilaç provizyon sisteminden alınan

solunum sistemi hastalıkları için yazılan antibiyotik veri seti üzerinde CART algoritması ve Lojistik Regresyon ile ortaya konmuştur. Sonuç olarak CART algoritmasının Lojistik Regresyona göre daha iyi bir sınıflama yaptığını gözlemlemiştir. Pehlivan (2006) yılında yaptığı araştırmada kurumsal bir firmada çalışan personelin eğitim, işe alma, performans, ücretlendirme, kariyer planlama özlük işleri, kurum içi iletişim, bordrolama ve bu firmanın insan kaynakları hakkındaki görüşlerini içeren anket formundan elde edilen veriler üzerine CHAID analizi uygulanarak bu firmada çalışanların özellikleri sınıflanmıştır.

Kayışoğlu (2014) akciğer kanseri hastaları üzerine yapmış olduğu çalışmada hastalığın doğru sınıflandırılmasının önemine vurgu yaparak Weka yazılımı ile doktorlara doğru ve süratle karar alabilmelerine yardımcı olabilecek bir sistem geliştirmiştir. Değirmenci (2014) TÜİK Hane Halkı Bütçe Araştırmasından (HBA) elde edilen veriler üzerine veri madenciliği kümeleme yöntemlerinden K-Ortalamalar tekniğini kullanarak resmi istatistikler üzerine veri madenciliği tekniklerinin uygulanması ile örüntülerin ortaya çıkarılabileceğini göstermiştir.



3. VERİ MADENCİLİĞİNDE KULLANILAN ALGORİTMALAR

Veri madenciliği yöntemleri 3 ana başlıkta ifade edilmektedir. Bunlar; Sınıflama, kümeleme ve birliktelik kurallarıdır.

3.1. Sınıflama Algoritmaları

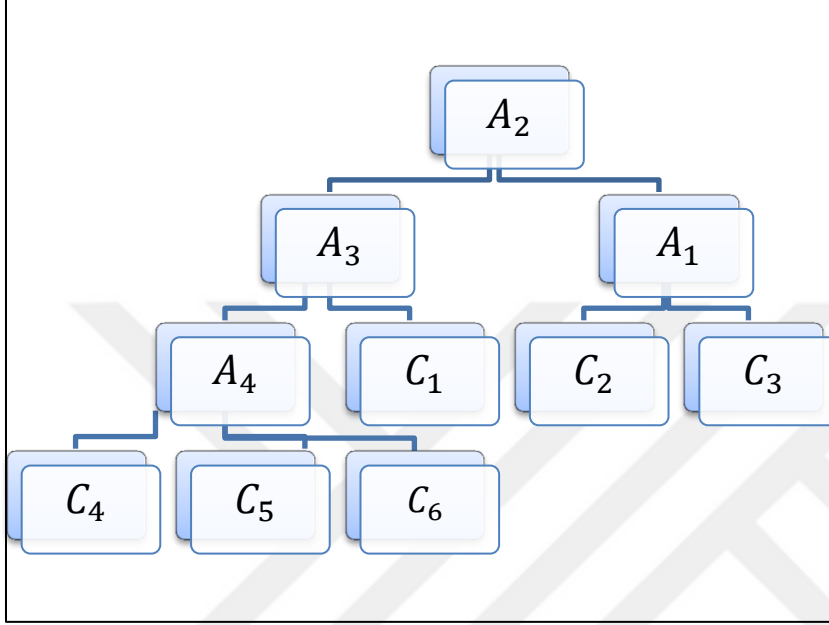
Sınıflama; Verideki bilginin ortak özelliklere göre ayrıştırılmasıdır. Sınıflamada izlenen süreç, mevcut veri tabanından yararlanılarak kuralların oluşturulması ve bu kurallar yardımıyla kararın verilmesidir (Özkan, 2008: 51).

Sınıflama uygulamalarına örnek olarak davranış tanımlama, tıbbi teşhis, endüstriyel hataların tespiti ve alış veriştir trendlerini sınıflama olarak verilebilir. Tahmin ve kestirim sınıflamanın bir çeşidi olarak görülebilir. Biri yaşımızı tahmin ettiğinde ya da bir kavanozun içerisindeki bilyeleri tahmin etmeye çalıştığında bunlar gerçekte sınıflama problemleridir. Sınıflamayı uygulayabilmek için geliştirilen tüm yaklaşımlar bazı verilerin bilinmesini varsaymaktadır. Sıklıkla bir kontrol grubu, kullanılan teknik için gerekli olan belirli parametreleri geliştirmek için kullanılmaktadır. Sınıflama veri tabanından sınıfların kümelerini haritalama olarak görülebilmektedir.

Birçok sınıflama yöntemi vardır, karar ağaçları, istatistiğe dayalı algoritmalar (Bayesyen ve Regresyon), mesafeye dayalı sınıflama algoritmaları (k-en yakın komşu), yapay sinir ağları bunlardan bazılarıdır.

Karar ağacı yaklaşımı, hedef fonksiyonlarını yaklaşık olarak hesaplamak için kullanılan bir yöntemdir. Ağaç görünümünde tanımlayıcı ve tahmin edici bir modeldir. Bu model, karar alıcıya karar alırken hangi faktörlerin göz önüne alınması ve her bir faktörün farklı çıktıları ile geçmişte nasıl ilişkili olduğunun belirlenmesi konularında yardımcı olur. Karar ağaçlarında pek çok test gerçekleştirilerek hedefi tahmin etmede en iyi sırayı bulmaya çalışılır. Her bir test karar ağacındaki dalları oluşturur ve bu dallar da diğer testlerin gerçekleşmesine neden olur. Bu durum, test işleminin bir yaprak düğümünde (leaf node) sonlanmasına kadar devam eder. Kökten hedef yaprağa kadar olan yol, hedefi sınıflandıran “kural” olarak adlandırılır. Kurallar “eğer-sonra” (if-then) yapısındadır (Emel ve Taşkın, 2014).

Karar ağaçları, sınıflamada en yaygın kullanılan yöntemlerden biridir. Bu yöntem, tahmin edici değişkenleri kullanarak model kurulması ve bu model sayesinde yeni değişkenlerin üyelik derecelerine göre farklı sınıflara ayrılması işlemidir (Gemici, 2012). Şekil 3.1 karar ağacı algoritmalarına örnek ağaç olarak verilmiştir.



Şekil 3.1. Örnek karar ağacı

3.1.1. ID3 algoritması

Diğer değişkenler içerisinde sınıflamada en ayırıcı özelliğe sahip değişkeni bulurken entropi kavramından yararlanılır. Entropi, eldeki bilgilerin sayısallaştırılmasıdır. Bir veri kümesi içindeki belirsizlik ve rasgeleliği ölçmek için kullanılır. Örneğin herkes aynı futbol takımını tutsaydı, herhangi bir kişiye tuttuğu takımı sorduğumuzda aldığımız yanıt bizi şaşırtmayacaktı, bu durumda entropi sıfır (0) olacaktı. Entropi 0-1 arasında değer alır. Bütün olasılıklar eşit olduğunda entropi maksimum değerine ulaşacaktır.

Matematiksel olarak;

$\langle p_1, p_2, \dots, p_n \rangle$ olasılıkları ifade ederse tüm olasılıkların toplamı 1 (bir) olmalıdır. $\sum_{i=1}^n (p_i) = 1$ bu durumda entropi aşağıdaki gibi olacaktır.

$$H(p_1, p_2, \dots, p_n) = -\sum (p_i \log_2(p_i)) \quad (3.1)$$

ID3 algoritması bir karar ağacı algoritmasıdır. Karar ağacı şekil itibariyle akış şemasına benzemektedir. Karar ağacını oluşturan elemanlar; kök, düğüm, dal ve yaprak olarak isimlendirilir (Özkan 2008: 55, 75). Karar ağaçlarında algoritmalar dallanmanın nasıl yapıldığına göre farklılık göstermektedir. ID3 entropiye dayalı dallanma yapan bir algoritmadır. Entropi bir verideki belirsizliği ya da rastgeleliği ölçmekte kullanılmaktadır. Bir sınıfta bulunan tüm verilerde belirsizlik yoksa bu durumda entropi değeri sıfır olur. Karar ağacı algoritmalarının sınıflama amacı, aynı sınıfa ait tüm elemanların iteratif olarak alt setlere ayrılmasıdır (Dunham, 2003: 97).

Kök: Karar ağacında en üst yapıya kök denir.

Düğüm: Karar ağaçlarında her bir niteliğe düğüm denir.

Dal: Karar ağaçlarında kök ve yaprak arasında kalan yapılar dal olarak isimlendirilir.

Yaprak: Karar ağacı yapısında en son elemana yaprak denir.

Örnek:

Çizelge 3.1. ID3 algoritması için örnek olaylar ve belirsizlikler

Deney Sonuçları (S)	a_1	a_2	a_3
P_i	1/3	1/4	1/7

$S=\{ a_1, a_2, a_3 \}$ deney kümesinde a_1, a_2, a_3 olaylarının belirsizlikleri (entropileri) sırayla aşağıdaki gibidir;

$$-\frac{1}{3}\log_2 \frac{1}{3}$$

$$-\frac{1}{4}\log_2 \frac{1}{4}$$

$$-\frac{1}{7}\log_2 \frac{1}{7}$$

$$\text{Toplam Belirsizlik (Entropi)}= H(S)=-(\frac{1}{3}\log_2 \frac{1}{3} + \frac{1}{4}\log_2 \frac{1}{4} + \frac{1}{7}\log_2 \frac{1}{7})=1,4293$$

Karar ağacı oluştururken dallanmaya hangi nitelikten başlanacağı önemlidir. Bu niteliği belirlemek için kazanç (gain) adı verilen bir ölçüt kullanılmaktadır. ID3 Algoritmasının kullandığı kazanım şu şekilde hesaplanır;

Verilerin ham halinin (başlangıçtaki) entropisi ile her bir alt bölümün entropilerinin ağırlıklı toplamı arasındaki fark alınır. Bu fark hangi alt bölüm için büyükse o alt bölüme doğru dallanma yapılır.

$$\text{Kazanım (D;S)} = H(D) - \sum_{i=1}^n P(D_i)H(D_i) \quad (3.2)$$

Bu ölçüte göre en yüksek kazancı veren nitelik kök düğümü oluşturur ve benzer şekilde iterasyonlar uygulanarak sonraki düğümler belirlenerek karar ağacı oluşturulur (Dunham, 2003: 98).

3.1.2. C4.5 algoritması

C4.5 algoritması ID3 algoritmasının genişletilmiş halidir. Karar ağacı oluştururken optimal dallanmayı belirlemek için ID3 algoritmasına benzer şekilde kazanım ve entropi değerleri kullanılmaktadır. C4.5 algoritması kayıp verileri diğer veri ve değişkenler yardımıyla öngörerek kazanım oranının hesaplanmasında kullanılır. Böylece daha duyarlı ve daha anlamlı kurallar çıkartabilen bir ağaç üretebilir. Öncelikle en yüksek bilgi sağlayan t eşik değerinin belirlenmesi için nitelik değerler sıralanır $\{v_1, v_2, \dots, v_n\}$, $t_i = (v_i + v_{i+1})/2$ şeklinde orta nokta belirlenir ve nitelik için t ile karşılaştırmalar yapılır (Özkan 2008: 77). Karşılaştırma sonucunda dallanmanın nereden yapılacağına karar verilir.

C4.5 Algoritması, ID3 Algoritmasını geliştirilmesi ile elde edilmiştir. C4.5 algoritmasının ID 3 algoritmasına üstün olduğu konular şu şekildedir;

Kayıp değerler: Karar ağacı oluştururken kayıp değerler basitçe göz ardı edilebilmektedir. Yani kazanım oranı hesaplanırken sadece verileri eksik olmayan diğer kayıtlar kullanılır.

Budama: C4.5 Ağacın budanması işlemi için iki ayrı yöntem kullanılır.

Alt ağaç yerleştirmesi: Ağaç içindeki alt-ağaçlar yapraklara dönüştürülür. Ancak bu değişimin yapılabilmesi için konulacak yaprağın hatasının, alt ağacın hatasından düşük

olması gerekir. İkinci yöntem ise bir alt-ağacın, bu ağacı en çok kullanan ağacın yerini almasıdır. Yine burada da hata oranlarının yapılan değişiklikten sonra düşmesi gerekmektedir.

Sayısal Değerler: Sayısal değere sahip değişkenlerle karar ağacı oluşturmaya olanak sağlamaktadır.

3.1.3. CART algoritması

Breiman ve diğerleri tarafından 1984'te önerilen CART tekniği ID3 algoritmasında olduğu gibi dallara ayırmada optimal kriteri seçmek için entropiden yararlanır. Ancak en optimal kriteri belirlemek için ID3 ve C4.5 ten farklı bir formül kullanılır. CART algoritmasında kullanılan optimal dallandırma kriteri şu şekilde hesaplanır;

CART algoritması tarafından üretilen karar ağaçları keskin bir şekilde ikili olarak dallanmaktadır. Herbir düğümde ikili olarak dallanmaktadır. CART dallara ayırma kriteri hesaplanırken kayıp veriler önemsenmez. CART algoritmasında entropi değeri aşağıdaki şekilde hesaplanmaktadır (Larose, 2005: 109, 110).

Tüm niteliklerin var olduğu değerler göz önüne alınır ve eşleşmelerden sonra twoing ve gini algoritması olarak 2 bölünme elde edilir.

Twoing algoritmasında; aday bölünme denilen sağ-sol olarak 2 dal bulunur ve her biri için olasılıklar hesaplanır.

Herhangi bir t düğümündeki s dallara ayrılma kriteri $\Phi(s/t)$ olarak gösterilirse;

$$\Phi(s/t) = 2P_L P_R \sum_{j=1}^{sınıflar} |P(j|t_L) - P(j|t_R)| \quad (3.3)$$

t: Dalların yapılacağı düğüm

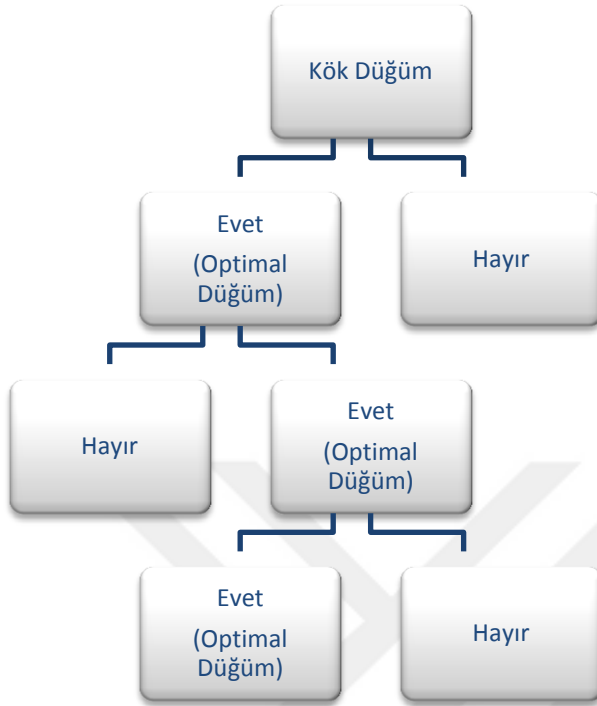
C: Kriter

L: Ağacın sol tarafı

R: Ağacın sağ tarafı

P_L, P_R : Öğrenim kümesindeki bir kaydın sağda ve solda olma olasılıkları

$P(J|t_L)$ ve $P(J|t_R)$: C_J sınıfındaki bir kaydın sağda ve solda olma olasılıkları



Şekil 3.2. CART algoritması karar ağacı

CART algoritması sonucunda elde edilen karar ağacının ikili dallanma yaptığı Şekil 3.2’de gösterilmektedir.

3.1.4. SPRINT algoritması

Karar ağacı tabanlı ilk paralel sınıflama algoritmalarındandır. Bu ölçeklenebilir algoritma çok sayıda işlemcinin paralel olarak bir karar ağacını beraber üretmesine dayanır. SPRINT algoritmasında girdilerin boyutu için herhangi bir sınırlama bulunmaz. Sonuca hızlı ulaşabilen bir algoritmadır. Veri kayıtları ile veri değerleri, bir sınıf etiketi ve bu değerlerle elde edilen kayıt endeksi oluşur. Bu kayıt endeksi “rid” olarak gösterilir. İlk tablolar sürekli değişkenlere göre sıralanır. Eğitim verilerinden elde edilen ilk tablolar, karar ağacının kök düğümüyle ilişkilendirilir. Ağaç bu şekilde gelişir ve bölünmelerle yeni ürünler oluşur ve her düğüme ait değişken tabloları bölünür ve bu ürünlerle ilişkilendirilir. Bir tablo bölündüğünde kayıtların sırası korunur. Bundan dolayı, tekrar sıralama yapmaya gerek kalmaz. Ayırma şartları altında sürekli değişkenler için her karar ağacı düğümü iki histogramla ilişkilendirilir. Bu histogramlar, ALT C ve ÜST C ile gösterilir ve bu histogramlar kayıtların sınıf dağılımlarını belirler. Kategorik değişkenler, bir düğüme

ilişkili olarak histogramı çıkarılır. Bu tek histogram, her değişkenin sınıf dağılımını gösterir. Bu histograma sayı matrisi adı verilir (Shafer ve diğerleri, 1996:546). Sprint algoritmasında düğümlerden alt dallara ayırma kriteri için gini(K) endeksi kullanılmaktadır.

$$gini(K)=1-\sum p_j^2 \quad (3.4)$$

p_j : K kümesi içindeki j düzeyinin olasılığıdır.

Eğer K kümesi K_1 ve K_2 gibi alt kümelere ayrılırsa bölünmüş k kümesinin $gini_{bölünmüş}(K)$ değeri;

$$gini_{bölünmüş}(K) = \frac{n_1}{n} gini(K_1) + \frac{n_2}{n} gini(K_2) \quad (3.5)$$

şeklinde hesaplanmaktadır (Silahtaroglu, 2008: 59,60).

3.1.5. Bayesyen sınıflama

Bayesyen sınıflama elde mevcut olan sınıflanmış veriler kullanılarak yeni verinin var olan sınıflardan birine girme olasılığını hesaplayan bir tekniktir. Bayesyen sınıflama bilim insanlarına deneyimlerinin ya da var olan bilgilerle, yeni bilgileri birleştirilmesine yardımcı olmaktadır. Bir öğrenme seti kullanan bayes sınıflaması, tahmin edici değişkenlerin aldığı değerlerin belirli uzaklıklarına göre ilişkili oldukları sınıfların olasılıklarını hesaplamak için Bayes teoremine dayanan bir metottur. (Dumitru, 2009).

Bayes teoremine göre kullanılacak formül;

$$P(C_1 \setminus x_i) = \frac{P(x_i \setminus C_1)P(C_1)}{P(x_i \setminus C_1)P(C_1) + P(x_i \setminus C_2)P(C_2)} \quad (3.6)$$

C_1 ve C_2 olarak gösterilen iki ayrı sınıfın olduğu kabul edilmiştir.

$P(C_1 \setminus x_i)$ x_i 'nin C_1 sınıfında olma olasılığını ifade etmektedir.

$P(x_i)$, x_i değerinin veri tabanında bulunma olasılığıdır.

n sayıda sınıf olduğu varsayılırsa bu durumda kural;

$$P(x_i) = \sum_{j=1}^n P(x_i \setminus C_j)P(C_j), \quad j = 1, \dots, n \quad (3.7)$$

Buna göre x_i 'nin C_1 sınıfında bulunma olasılığı;

$$P(C_1|x_i) = \frac{P(x_i|C_1)P(C_1)}{P(x_i)} \quad (3.8)$$

formülü ile hesaplanmaktadır (Silahtaroglu 2008:61).

3.1.6. K-En yakın komşu

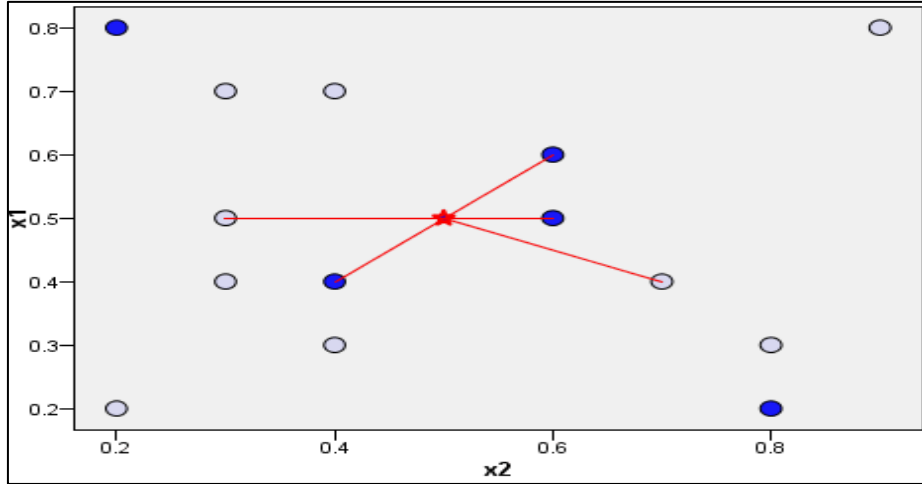
En çok kullanılan algoritmalarından bir diğeridir. Kaynaklarda K-Nearest Neighbor (KNN) şeklinde geçmektedir. Üzerinde çalışılacak veri setindeki her bir kaydın diğer kayıtlara olan mesafeleri hesaplanmaktadır. Ancak bunu yaparken diğer kayıtlardan k sayıda olan kayıt dikkate alınmaktadır. Veri tabanındaki bu k sayıda nokta uzaklığı hesaplanan noktaya diğer noktalara göre en yakın olan kayıtlardır. Coğrafi bilgi sistemlerinde kullanılan bu yöntem arzu edilen bir noktaya en yakın şehir, bölge, istasyon belirlenmesi k-en yakın komşu algoritmasını oluşturmaktadır (Beyer, Goldstein, Ramakrishnan ve Shaft, 1999).

Önce k değeri belirlenir. K değerinin büyük olması birbirlerine benzemeyen değerlerin bir araya gelmesine, düşük seçilmesi durumunda ise birbirine benzediği, aynı sınıfın verileri olmalarına rağmen bazı noktaların ayrı sınıflara düşmesine ya da apayrı yeni sınıfların oluşmasına neden olmaktadır. Bu nedenlerden dolayı çalışmalarda yoğunlukla kullanılan tipik k değerleri 3, 5 ve 7 şeklindedir (Khan, Ding ve Perrizo, 2002).

Yaygın kullanılan benzerlik fonksiyonlardan biri iki nokta arasındaki uzaklığı dayanan Öklid uzaklığıdır (Kabak, Sağlam ve Aktaş, 2017). $X = x_1, x_2, x_3, \dots, x_{n-1}$ ve $Y = y_1, y_2, y_3, \dots, y_{n-1}$ olduğunda X ve Y noktaları arasındaki doğrunun Öklid uzaklığı;

$$d(X, Y) = \sqrt{\sum_{i=1}^{n-1} (x_i - y_i)^2} \quad (3.9)$$

formülü ile hesaplanmaktadır. Kümeleme analizinde uzaklık ölçümü olarak ayrıca Manhattan veya Minkowski uzaklık ölçüleri sayılabilir. Bir noktaya olan en yakın komşuluklar Şekil 3.3'de gösterilmektedir.



Şekil 3.3. $K = 5$ olarak seçildiğinde belirlenen noktaya en yakın komşular

3.1.7. Yapay sinir ağları

Yapay sinir ağları, sinir hücrelerinin (nöron) birbirlerine farklı şekilde bağlanmalarıyla oluşmaktadır. Genel olarak tabakalar halinde düzenlenir. Her bir sinir hücresinde girdileri tanımlayan bir fonksiyon vardır. Her bir nöron diğer nöronlara biyolojik sinir hücrelerinde olduğu gibi sinyal göndermektedir. Sinir ağı içerisinde yer alan bir nöron birden fazla diğer nörona aynı anda sinyal gönderebilmektedir (Fausett, 1994:3).



Resim 3.1. Yapay sinir ağı

Yapay sinir ağıları, sınıflamanın yanında kümeleme analizi ve örüntü tanımlama işlemlerinde de kullanılabilir. Sınıflama açısından belirlenen öğrenmeyi (kural) algoritma sürdürür ve çıktıya ulaşmak için ağırlıklar hesaplanır. Öğrenme bir kez tamamlandıktan sonra yeni verinin ağırlıklarının hesaplanması kolay olmaktadır. Yapay sinir ağlarının öğrenilmesi çok uğraştırıcıdır ancak sonrasında çok duyarlı sınıflamalar yaptığı görülmektedir (Silahtaroglu, 2008).

3.2. Kümeleme Algoritmaları

Kümeleme; veriler arasındaki benzerlikler değerlendirilerek bunların gruplandırılmasına denir, pazarlama araştırmalarında yaygın olarak kullanılmaktadır. Kümeleme hiyerarşik ve hiyerarşik olmayan kümeleme olmak üzere ikiye ayrılmaktadır. Hiyerarşik kümelemeye en yakın komşu ve en uzak komşu algoritması örnek verilebilir. Hiyerarşik olmayan kümelemeye ise k-ortalamlar yöntemi örnek verilebilir. Kümeleme, verilerin gruplanması açısından sınıflamaya benzerdir. Ancak sınıflamanın aksine gruplar önceden tanımlanmamaktadır. Bunun yerine gruplama, gerçek veride bulunan karakteristiklere göre veriler arasındaki benzerlikler bulunarak gerçekleştirilmektedir. Gruplar küme olarak adlandırılmaktadır. Bazı yazarlar kümelemeyi sınıflamanın bir çeşidi olarak görmektedirler. Kümelemeye benzer bir terim de, veri tabanındaki kayıtların birlikte gruplanması şeklinde ifade edilebilecek veri tabanının parçalara ayrılmasıdır (database segmentation). Bu terim, araştırmacıya veri hakkında daha genel bir görüş verecek bileşenlerin veri tabanı içinde parçalara ayrılması şeklinde tanımlanabilmektedir. Kümeleme, biyoloji, tıp, antropoloji, piyasa ve ekonomi alanlarında kullanılmaktadır.

3.2.1. En yakın komşu algoritması (SLINK)

SLINK algoritmada iki küme arasındaki mesafe ölçülürken kümeler içinde birbirlerine en yakın iki gözlemin uzaklığı kümeler arasındaki uzaklık olarak kabul edilir. Bu yöntemde öncelikle gözlemler arasındaki uzaklıklar hesaplanır (Sibson, 1973). Hesaplama yapılırken Öklid uzaklık bağıntısı kullanılabilir Buna göre i ve j noktaları için uzaklık ölçüsü;

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3.10)$$

şeklinde hesaplanır. Hesaplanan uzaklıklar arasında $\min d(i,j)$ uzaklığı en yakın iki kümenin uzaklığını vermektedir.

Tüm uzaklıklar belirlendikten sonra benzerlik/uzaklık/mesafe tablosu oluşturularak benzer gözlemler bir araya getirilerek kümeler oluşturulur. Kümeler içerisine girecek yeni veriler en yakın mesafe/benzerliğe göre belirlenmektedir. Bu yöntemin en karmaşık yanı gözlem sayısının çokluğuna bağlıdır. Her bir döngü için her bir gözlem değeri kümede zaten var olan verilerle karşılaştırılmak zorundadır. Bu da oldukça zaman almaktadır. Yöntemin en dezavantajlı yanı budur (Dunham, 2002:142).

3.2.2. En uzak komşu algoritması (Cure)

En yakın komşu algoritmasına benzeyen yöntem, SLINK algoritmasının aksine bu defa iki kümenin elemanları içerisinde birbirine en uzak olanlar arasındaki mesafe kümeler arasındaki uzunluk olarak belirlenmektedir. Yine uzaklık ölçüsü olarak Öklid bağıntısı kullanılabilir.

Uç değerlere göre hassas olan diğer kümeleme algoritmalarına göre Cure algoritması uç değerlere karşı daha güçlü bir yapıya sahiptir. Küresel olmayan ve çok geniş varyanslara sahip kümeleri tanımlayabilmektedir. Cure büyük verilerle çalışabilmek için öncelikle veri setinden örnek veriler alarak bunları parçalara ayırdıktan sonra kümeleme yapmaktadır (Guha, Rastogi ve Shim, 1998).

3.2.3. CHAMELEON algoritması

Diğer algoritmalarından farklı şekilde iki alt kümenin benzerlikleri ve yakınlıklarının kıyaslandığı homojen ve doğal dinamik bir model kullanmaktadır. Kümeleme analizinde çoğu algoritma kümeler farklı şekillerden, sıklıklardan ve büyüklüklerden oluştuğunda hatalı çalışmaktadır. CHAMELEON , Karypis, Han ve Kumar (1999) tarafından iki kümenin benzerliğini dinamik bir model ile ölçen algoritma olarak geliştirilmiştir. Bu dinamik modele dayanan CHAMELEON algoritması, benzerlik matrisi kurulabildiği sürece tüm veri çeşitleri için uygulanabilmektedir. Kümeleme sürecinde iki küme, alt kümelerin kendi iç benzerlikleri ve yakınlıkları karşılaştırılarak yakın ise birleştirilmektedir. Bu algoritmanın en önemli özelliği kümeler birleştirilirken yalnızca

kümelerin birbirlerine olan yakınlıklarına değil aynı anda her bir kümenin iç benzerliğine ve yakınlığına yani küme içinde bulunan noktalar arasındaki bağlantıya bakmasıdır. C_j ve C_i kümeleri birleştirilirken hem göreceli iç bağlantısallık (relative inter-connectivity) $RI(C_j, C_i)$ hem de göreceli yakınlık (relative closeness) $RC(C_j, C_i)$ değerleri hesaplanır. $RI(C_j, C_i)$ ve $RC(C_j, C_i)$ değerleri yüksek hesaplanan kümeler birleştirilmektedir.

$$RI(C_j, C_i) = \frac{2 \times |EC_{(C_j, C_i)}|}{|EC_{C_i}| + |EC_{C_j}|} \quad (3.11)$$

$|EC_{(C_j, C_i)}|$, C_j ve C_i kümeleri arasındaki mutlak bağlantısallıktır. $|EC_{C_i}|$, C_i kümesinin kendi iç bağlantısallığı, $|EC_{C_j}|$ ise C_j kümesinin iç bağlantısallığıdır.

$$RC(C_j, C_i) = \frac{\bar{SEC}_{(C_j, C_i)}}{\frac{|C_i|}{|C_i + C_j|} \bar{SEC}_{C_i} + \frac{|C_j|}{|C_i + C_j|} \bar{SEC}_{C_j}} \quad (3.12)$$

$\bar{SEC}_{C_i}$ ve $\bar{SEC}_{C_j}$, C_j ve C_i kümelerinin kendi iç noktalarının bağlantı kenarlarının ağırlıklarının ortalamasıdır. $\bar{SEC}_{(C_j, C_i)}$ ise C_j ve C_i kümelerini birbirine bağlayan kenarların ağırlıklı ortalamasıdır (Karypis ve diğerleri, 1999).

3.2.4. K-Ortalama algoritması

K-ortalamlar kümelerle veri noktaları ile kümeler arasındaki ilişki hakkında ön bilgi gerektirmeyen yaygın kullanılan bir kümeleme algoritmasıdır. K-ortalamlar algoritması, uzaklık ölçüsüne göre veri noktalarını k kümede gruplandırmaktadır (Orhan, Hekim ve Özer, 2011). K ortalamlar kümelerin devamlı yenilendiği, optimum çözüme ulaşana kadar devam eden döngüsel bir algoritmadır. Eldeki verilerle araştırmacı tarafından belirlenen k sayıda kümenin ortalamlarına göre ayrılması şeklinde hesaplanmaktadır. Burada ortalama küme merkezini ifade etmektedir. Yalnızca sayısal verilerle çalışabilen bu algoritmada kategorik veriler kullanılmamaktadır. K ortalamlar algoritmasının durma kuralı üst üste kümelerde aynı değerin bulunmasıdır (Macqueen, 1967).

$k_i = \{t_1, t_2, \dots, t_m\}$ kümesinin ortalaması;

$$m_i = \frac{1}{m} \sum_{i=1}^m t_i \quad (3.13)$$

şeklinde tanımlanır (Dunham, 2002:140).

K-ortalamalar örnek üzerinde anlatılacak olursa $D=\{4, 5, 8, 13, 15, 20\}$ eldeki veriler olduğunda ve bu altı noktanın $k=2$ kümeye ayrılması istendiğinde;

Keyfi olarak m_1 ve m_2 değeri belirlenir. Bu değerler D kümesinden $m_1 = 5$ ve $m_2 = 15$ olarak seçilsin. Bu durumda 4 ve 8 $m_1 = 5$ değerine yakın olduğundan k_1 kümesine, 13 ve 20 $m_2 = 15$ değerine yakın olduğundan k_2 kümesine dahil olacaklardır. Buradan yeni m_1 ve m_2 değerleri hesaplanarak $m_1 = 6$ ve $m_2 = 16,5$ olarak hesaplanır ve yeni m_1, m_2 değerlerine göre;

$$k_1 = \{4, 5, 8\}$$

$$k_2 = \{13, 15, 20\}$$

Şeklinde kümelenecektir. Bu iterasyonlar serisi, durdurma kuralı olan bir önceki iterasyonda hesaplanan k_1 ve k_2 kümelerinin elemanları aynı oluncaya kadar devam etmektedir.

3.2.5. PAM algoritması

Güçlü matematiksel ve istatistiksel geçmişi olmayan gene kullanıcılar için bir çok kümeleme metodunun anlatıldığı eserlerinde Kaufman ve Rousseeuw (2005) PAM algoritmasını geliştirmişlerdir. K-medoids adı da verilen PAM algoritmasının temsilci olarak belirlediği noktalara medoid denilmektedir. Diğer kümeleme metodlarına göre k-medoids var olan aykırı değerlere karşı oldukça güçlü bir yaklaşımdır. Ayrıca veri yapısındaki ortogonal dönüşümlere ve veri dönüşümleri açısından değişmezlik özelliğine sahiptir. Yapılan birçok deneyde büyük veri setleri ele alındığında k-medoids algoritmasının etkinliği gösterilmiştir. PAM algoritması, her bir küme için bir temsilci belirleme yaklaşımına sahiptir. Bu temsilcilere medoid denildiği yukarıda belirtilmiştir. Medoidler seçildiğinde, geriye kalan seçilmemiş her bir nesne en çok benzediği medoidler ile gruplanmaktadır.

p_i bir medoid ve p_j bir medoid olmayan nokta olduğunda $d(p_i, p_j) = \min_{p_e} d(p_j, p_e)$ ise p_j , p_i 'nin temsil ettiği kümeye aittir. $\min_{p_e}$ tüm medoidler içerisinde en küçük olanı temsil etmektedir. $d(p_a, p_b)$ ise p_a ve p_b noktaları arasındaki uzaklığı ifade etmektedir. K-medoids algoritmasında tüm yakınlık ölçüleri kullanılabilir (Raymond ve Han, 1994).

3.3. Birliktelik Kuralları

Birliktelik kuralları; olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemidir. Bir ayakkabıcıdan ayakkabı satan alan birinin aynı mağazadan ayakkabı boyası da satın alması birlikte gerçekleşme durumuna bir örnektir. Bu algoritma en çok müşterilerin satın alma eğilimlerini tespit etmek amacıyla kullanılmaktadır. Satın alınan bir ürünle beraber bir başka ürün satın almak birliktelik kuralını temsil etmektedir. Birliktelik kuralları çoğunlukla perakende satış yapan mağazalarda satışa yardımcı olma veya reklam aracı olarak kullanılmaktadır. Perakende sektöründe direkt uygulanabilirliğe sahip olmasına rağmen, iletişim ağlarındaki hataların tahmin edilmesi gibi diğer amaçlarda da kullanılmaktadır. Birliktelik kuralları ürün verileri arasındaki ilişkileri göstermede kullanılmaktadır. Birliktelik kuralının bulunduğu bir veri tabanı, her birinin ürün setlerinden oluştuğu değişkenler demeti olarak görülebilir. Örneğin üç nesneden oluşan bir değişkenler demeti; tereyağı, ekmek ve reçelden oluşabilmektedir. Birliktelik kurallarında sayılacak algoritmalara AIS Algoritması, SETM Algoritması, Apriori algoritması, Apriori Tid Algoritması örnek olarak verilebilmektedir (Silahtaroglu 2008: 45, 83, 99). Birliktelik kurallarına dair bazı algoritmalar aşağıda açıklanmıştır.

3.3.1. Apriori algoritması

Apriori algoritması birliktelik kuralları içerisinde en bilinen ve en çok kullanılan algoritmalarından biridir. Apriori algoritmasının temel yaklaşımı büyük veriden, belirli büyüklükte bir aday set ürettikten sonra tüm veride tarama yaparak ne kadar büyük olduğunu görmek için bu aday seti oluşturan bilgileri saymaktır. Apriori veri tabanındaki değerler arasında bulunan önemli ilişkileri üretmektedir (Agrawal, Imielinski ve Swami, 1993).

Birliktelik kuralları “destek” ve “güven” olarak isimlendirilen iki ölçüte göre oluşmaktadır. Bu ölçütler hesaplanırken destek sayısı adında başka bir değer kullanılmaktadır. Destek, bir ilişkinin yapılan alışverişler arasında oransal olarak nasıl tekrarlandığını belirleyen ölçüttür. Güven ise A ürün grubunu satın alan kişilerin B ürün grubunu da satın alma olasılıklarını veren bir ölçüttür. Birliktelik kuralı $A \rightarrow B$ şeklinde ifade edilmektedir. Destek ve Güven ölçütleri için hesaplama formülleri aşağıdaki gibidir.

$$destek(A \rightarrow B) = \frac{sayı(A,B)}{N} \quad (3.14)$$

$$güven(A \rightarrow B) = \frac{sayı(A,B)}{sayı(A)} \quad (3.15)$$

$sayı(A,B)$, destek sayısını yani A ve B ürün gruplarının birlikte satın alındığı sayıyı göstermektedir. N ise yapılan tüm alışverişlerin sayısıdır (Tan, Steinbach ve Kumar, 2006:71,72).

Birliktelik kuralları için hesaplanan destek ve güven ölçütleri hesaplandıktan sonra bu değerleri karşılaştırmak için bir eşik değere ihtiyaç duyulmaktadır. Hesaplanan güven ve destek ölçütlerinin belirlenen bu eşik değerden yüksek olması beklenmektedir. Birliktelik kuralının güçlü olduğuna karar vermek için hesaplanan destek veya güven ölçütlerinin büyük olması gerekmektedir (Özkan, 2008:158).

3.3.2. AIS algoritması

Agrawal ve diğerleri tarafından 1993 yılında ortaya atılan AIS algoritması, büyük nesne kümeleri ortaya çıkarmak için geliştirilmiş bir algoritmadır. Veri içerisindeki kayıtlı nesnelerin adlarının A'dan Z'ye sıralanması kısıtına sahiptir. AIS algoritması veri tabanını birden fazla defa tarar ve her tarama işlemi sırasında tüm işlemleri okumaktadır. İlk etapta veri tabanında kayıtlı tüm noktalar birer birer sayılar olarak “geniş nesneler” olduğunu belirledikten sonra bunlar aday nesne kümeleri olarak belirlenmektedir. Bir işlem taraması yapıldığında, bir önceki taramada belirlenen geniş nesne kümeleri arasında ortak nesne kümeleri belirlenmektedir. Belirlenen bu ortak nesne kümeleri işlemdeki diğer nesnelerle birleştirilir ve yeni aday nesneler oluşturulur. Herhangi bir nesnenin aday kümelere biri olabilmesi için, birleşeceği nesnenin geniş nesne kümesi ve harf sırası bakımından tüm nesnelerden sonra geliyor olması gerekmektedir. AIS bu yönüyle iteratif bir algoritma

özelliđi taşımaktadır. AIS algoritması bu adımları gerçekleştirebilmek için aday kümeler içerisinden gereksiz olanların silindiđi bir budama tekniđine sahiptir. Sonrasında aday kümesinin destek deđeri hesaplanarak, minimum destek seviyesine eşit ya da daha büyük olanlar aday geniş nesne kümesi olarak belirlenirler. Bu geniş nesne kümeleri bir sonraki iterasyonda aday kümenin belirlenmesinde kullanılmaktadır (Agrawal ve diđerleri, 1993).



4. TÜRKİYEDE ANLAŞMALI BOŞANMALAR

Anlaşmalı boşanmalar sorunu, Türkiye sosyal güvenlik mevzuatında yer alan boşluğun sigortalılarca suiistimal edilmesi sonucu ortaya çıkmaktadır. Mevzuatta yer alan mevcut hükümlere göre, yetim kalan çocukları sosyal güvence çatısı altına almayı hedeflemesine rağmen, evlendikten sonra boşanan ileri yaştaki kız çocukları da yetim olarak nitelendirildiklerinden babadan dolayı aylığa müstahak olmaktadır. Ancak Avrupa birliği ve dünya sosyal güvenlik uygulamaları incelendiğinde yetim aylıklarının, erkek yetim veya kız yetim olduğu gözetilmeksizin 25 yaşında kesilmesi öngörmüştür (Ekdemir, 2005:105). Burada ülkemiz sosyal güvenlik mevzuatı yetim kız çocukları için dünya uygulamalarından pozitif ayrılmaktadır. Bunun sonucunda anlaşmalı boşanmaların ortaya çıkma durumu arasında ülkemizde aylığa hak kazanma şartlarında erkek ve kız çocukları için dünyadaki örneklerin aksine norm ve standart birliğin bulunmaması sayılabilmektedir. Bu kişilerin tespiti ise Sosyal Güvenlik Kurumunun denetim birimlerinin görev alanına girmektedir. Anlaşmalı boşanan kişilerin denetimi özel hayatın gizliliği ilkesi dikkate alındığında ve ayrıca fiili denetimden Kamu Kurumları ya da özel sektörle yapılan yazışmalara kadar çeşitli bürokratik süreçler gerektirmesi nedeniyle tespiti zor ve uzun bir konudur.

4.1. Ölüm Aylığına Hak Kazanma Şartları

5510 sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası kanunun 34 üncü maddesinde ölüm aylığının hak sahiplerine paylaştırılması, “ölen sigortalının 33 üncü madde hükümlerine göre hesaplanacak aylığının;

- a) Dul eşine % 50'si; aylık bağlanmış çocuğu bulunmayan dul eşine ise bu Kanunun 5 inci maddesinin birinci fıkrasının (a), (b) ve (e) bentleri hariç bu Kanun kapsamında veya yabancı bir ülke mevzuatı kapsamında çalışmaması veya kendi sigortalılığı nedeniyle gelir veya aylık bağlanmamış olması halinde % 75'i,
- b) Bu Kanunun 5 inci maddesinin birinci fıkrasının (a), (b) ve (e) bentleri hariç bu Kanun kapsamında veya yabancı bir ülke mevzuatı kapsamında çalışmayan veya kendi sigortalılığı nedeniyle gelir veya aylık bağlanmamış çocuklardan;
- 1) 18 yaşını, lise ve dengi öğrenim görmesi halinde 20 yaşını, yüksek öğrenim yapması halinde 25 yaşını doldurmayanların veya,

- 2) Kurum Sağlık Kurulu kararı ile çalışma gücünü en az % 60 oranında yitirip malûl olduğu anlaşılanların veya,
- 3) Yaşları ne olursa olsun evli olmayan, evli olmakla beraber sonradan boşanan veya dul kalan kızlarının, her birine % 25'i,
- c) (b) bendinde belirtilen çocuklardan sigortalının ölümü ile anasız ve babasız kalan veya sonradan bu duruma düşenlerle, ana ve babaları arasında evlilik bağı bulunmayan veya sigortalının ölümü tarihinde evlilik bağı bulunmakla beraber ana veya babaları sonradan evlenenler ile kendisinden başka aylık alan hak sahibi bulunmayanların her birine % 50'si, bağlanacağı” şeklinde hükmedilmiştir.¹

Ölüm aylığının yetim kız çocuklarına paylaştırılması, 34 üncü maddenin birinci fıkranın (b) bendinin (3) numaralı alt bendinde “Yaşları ne olursa olsun evli olmayan, evli olmakla beraber sonradan boşanan veya dul kalan kızlara” ödeneceği hüküm altına alınmıştır. Türkiyede anlaşmalı boşanmaya konu olan kişiler sosyal güvenlik kanununun bu hükmü kapsamında girenlerden oluşmaktadır.

Bağlanan aylıklar daha sonra hak sahibi niteliğinin ortadan kalktığı ayı takip eden aydan itibaren kesilmektedir. Yani aylık alan yetim kız çocuklarının evlenmesi halinde, ölen ana veya babasından dolayı bağlanan yetim aylığı kesilmektedir.

4.2. Anlaşmalı Boşanma

Türkiye’de evlendiği halde ölen ana veya babasının aylığını almak isteyen yetim kız çocukları ya da ölen eşinden de ölüm aylığı almak isteyen kadın eşler, evli oldukları eşleriyle anlaşmalı olarak boşanarak aynı evde yaşamaktadırlar. 5510 sayılı kanunun gelir ve aylık bağlanmayacak haller başlıklı 56 ıncı maddesinin son fıkrasında “Eşinden boşandığı halde, boşandığı eşiyle fiilen birlikte yaşadığı belirlenen eş ve çocukların, bağlanmış olan gelir ve aylıkları kesilir. Bu kişilere ödenmiş olan tutarlar, 96 ncı madde hükümlerine göre geri alınır” hükmü yer almaktadır. Bu hükme göre boşandığı halde eşiyle fiilen birlikte yaşayan yetim kız çocuklarının aylıklarının kesileceği ve bu durumda ödenmiş aylıkların ise faizi ile birlikte geri alınacağı hüküm altına alınmıştır. Ayrıca ödenmiş aylıkların tahsiline ilişkin Sosyal Güvenlik Kurumu’nun 27/09/2008 tarihli ve 27010 sayılı resmi gazetede yayımlanan Fazla veya Yersiz Ödemelerin Tahsiline İlişkin

¹ 5510 sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu

Usul ve Esaslar Hakkında Yönetmeliğin 5 inci maddesinin birinci fıkrasının (d) bendinde eşinden boşandığı halde eşiyle fiilen birlikte yaşadığı tespit edilen eş ve çocuklar hakkında, aynı maddenin üçüncü fıkrası gereğince Cumhuriyet Savcılığına suç duyurusunda bulunulacağı hüküm altına alınmıştır. Ancak bu kişiler hakkında açılan davalar neticesinde, mahkemeler söz konusu durumun suç olmadığına dair hüküm vermiştir. Bunun sonucunda Sosyal Güvenlik Kurumu Fazla veya Yersiz Ödemelerin Tahsiline İlişkin Usul ve Esaslar Hakkında Yönetmelikte giderek bu kişiler hakkında suç duyurusunda bulunulmamasına dair değişiklik yapmış ve bu değişiklik 04/05/2013 tarihli ve 28637 sayılı Resmi Gazete’de yayımlanarak yürürlüğe girmiştir.²

4.3. Anlaşmalı Boşanmanın Tespiti

Anlaşmalı boşanmaların tespiti için denetimle görevli personellerin yapması gerekenler Sosyal güvenlik Kurumunca 29 Ocak 2013 tarihli ve 2013/5 sayılı Genelge ekinde yer alan Denetim Standartları ile talimatlandırılmıştır. Bu talimatlar;

- Kamu kurumları ve kolluk kuvvetlerinden alınan tek belge ve ifade ile yetinilmeyerek çevre incelemesi yapılması,
- Denetim esnasında kişilerin boşandığı eşi ile birlikte yaşadığını beyan eden ancak yazılı ifade vermek istemeyen kişilerin, temin edilebiliyorsa isim ve adreslerinin tespit edilmesi ve tutanağa yazılması,
- Boşandığı eşi ile birlikte yaşadığı iddia edilen kişiler üzerine bulunan aboneliklerin, ilgili kurumlardan yazı ile talep edilmesi ve araştırılması (Elektrik, su, doğalgaz),
- Boşandığı eşi ile birlikte yaşadığı iddia edilen kişiler kamu personeli ise, Kurumlarına yazı yazılarak izinde geçirdikleri adresler ve kurumlarına beyan ettikleri medeni durumlarının sorulması,
- Sağlık sisteminden boşandığı eşi ile birlikte yaşadığı iddia edilen kişilerin hastane kayıtlarının sorgulanması ve hastanelere verdikleri adres beyanlarının araştırılması,

şeklinde sıralanmıştır. Ancak bu hususta açılan davalar sonucunda mahkemeler, denetim sonucunda bir karara varılabilmesi için çok detaylı araştırmalar yapılması gerektiğine hükmetmektedir.

² Fazla veya Yersiz Ödemelerin Tahsiline İlişkin Usul ve Esaslar Hakkında Yönetmelik

Bu konuda açılan davalar sonucunda mahkemelerin araştırılmasını istediği hususlar ise yine denetim standartlarında mevzuat hükümleri göz önünde bulundurulmak suretiyle araştırma yapılması, tarafların göstereceği tüm kanıtların toplanması, bildirilen ve dinlenilmesi istenilen tanıkların ifadelerinin alınması, davacı ile boşandığı eşinin yerleşim yerlerinin saptanmasına ilişkin olarak; muhtarlıktan ikametgâh senetlerinin elde edilmesi, ilgili Nüfus Müdürlüklerinden sağlanan nüfus kayıt örnekleri ile yerleşim yeri ve diğer adres belgelerinden yararlanılması, adres değişiklik ve nakillerine ilişkin bilgilerin araştırılması, özellikle ilgili Nüfus Müdürlüğü'nden adres hareketleri tarihleriyle birlikte istenmesi, ilgililerin su, elektrik, telefon aboneliklerinin hangi adreste kimin adına tesis edildiğinin saptanması, seçmen bilgi kayıtlarının tespit ettirilmesi, tarafların varsa çalışmaları nedeniyle resmi/özel kurum ve kuruluşlara verilen belgelerde yer alan adreslerinin dikkate alınması, boşanan eşler 4857 sayılı Kanun hükümleri kapsamında yer almakta iseler adlarına ödeme yapılabilecek özel olarak açılan banka hesabı bulunup bulunmadığının belirlenmesi, boşanan eşlerin kayıtlı oldukları bölge/bölgeler yönünde kapsamlı Emniyet Müdürlüğü/Jandarma Komutanlığı araştırmasının yapılması, anılan mahalle/köy muhtar ve azalarının tanık sıfatıyla bilgi ve görgülerine başvurulması, gerektiği şeklinde sıralanmıştır.

Mahkemelerin verdiği kararlara bakıldığında bu konuda fiili tespitten, kurumlarla yapılacak bürokratik yazışmalara kadar çok detaylı yapılacak araştırmalar sonucunda “boşanılan eşle eylemli olarak birlikte yaşama” olgusunun gerçekleşip gerçekleşmediğine karar verilmesi gerektiğine hükmedilmektedir.

5. UYGULAMA

Ölen anne veya babasından aylık almak için anlaşmalı boşanarak eşiyle aynı evde yaşayan kişilerin tespitinin oldukça zor ve zaman alıcı olması Sosyal Güvenlik Kurumu için önemli bir problemdir. Bu nedenle, daha önce tespit edilen anlaşmalı boşanan grubun ortak özellikleri belirlenerek davranış modeli ortaya koymak araştırmanın temel amacını oluşturmaktadır. Bunun için Sosyal Güvenlik Kurumu veri tabanından anonim temin edilen verilere veri madenciliği uygulamalarında kullanılan SPSS Modeler 18. Programı ile veri madenciliği sınıflama tekniklerinden veri yapısına uygun olması ve uygulamalarda yaygın olarak tercih edilmesi nedenleriyle CART Algoritması kullanılarak aylıkları anlaşmalı boşanma kapsamında kesilen kişilerin Sosyal Güvenlik Kurumu veri tabanlarında tutulan karakteristik özellikleri sınıflanmıştır.

Araştırmanın uygulama bölümünde veri madenciliği alanında sık kullanılan paket programlardan biri olması, görsel programlama tekniklerine izin verdiği ve büyük verilerle çalışma konusunda elverişli olması nedenleriyle SPSS Modeler Programı kullanılmıştır. Büyük verilerle çalışmaya imkan sağlayarak veri madenciliğinde kullanılan diğer paket programlar KNİME, SAS, Weka şeklinde sıralanabilir (Kocabaş, 2010).

5.1. Uygulamada Kullanılan Değişkenlerin Tanımları

Araştırmada anlaşmalı boşandığı tespit edilen yetim kız çocukları üzerine veri madenciliği tekniklerinden karar ağacı algoritması uygulanacaktır. Bu algorithmada kullanılmak üzere Sosyal Güvenlik Kurumu veri tabanlarında anlaşmalı boşandığı tespit edilen kişiler hakkında anonim verilerden temin edilen değişkenler il, yaş, aylık türü, ölüm-boşanma süresi, prim ödeme gün sayısı, banka ve sektördür. Bu değişkenlere ilişkin bilgiler aşağıda özetlenmiştir.

İl: Anlaşmalı boşanan sigortalının ikamet ettiği ili ifade etmektedir. Türkiye genelinde ikamet eden bu kişiler için iller Türkiye'nin 7 bölgesine göre kategorize edilmiştir.

Çizelge 5.1. İllerin bulunduğu bölgelere göre belirlenen kategoriler

Bölgeler	Kategoriler
İç Anadolu Bölgesi	1
Karadeniz Bölgesi	2
Marmara Bölgesi	3
Ege Bölgesi	4
Akdeniz Bölgesi	5
Güneydoğu Anadolu Bölgesi	6
Doğu Anadolu Bölgesi	7

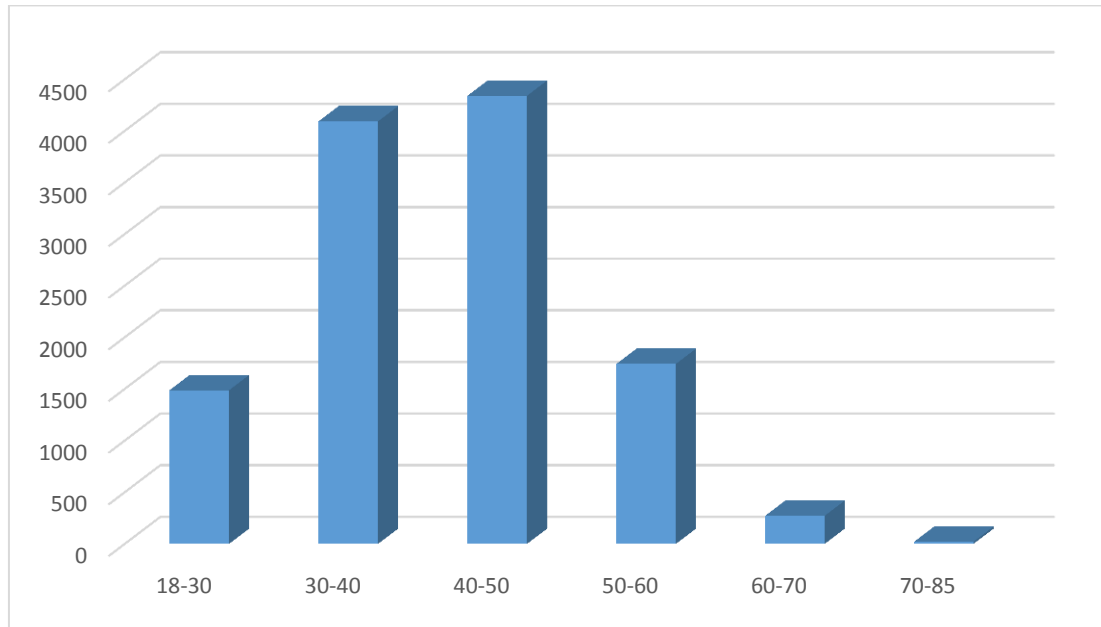


Harita 5.1. Türkiye'nin 7 bölgesinin gösteren harita

Yaş: Bu değişken Sosyal Güvenlik Kurumu veri tabanlarında tutulmamakta olup, kişilerin anlaşmalı boşanmayı gerçekleştirdiği yaş tespit edilirken veri tabanında tutulan aylık bağlama tarihi kullanılarak yaş değişkeni üretilmiştir. Yaş değişkeni, anlaşmalı boşanan kişinin ölüm aylığının veya ölüm gelirinin bağlandığı tarihteki yaşı ifade etmektedir. Uygulamada kullanılmak üzere 6 kategoride ifade edilerek, kategorik veriye dönüştürülmüştür.

Çizelge 5.2. Yaş aralıklarına göre kategoriler

Yaş Aralıkları(Yıl)	Kategoriler
17<Yaş≤30	1
30<Yaş≤ 40	2
40<Yaş≤ 50	3
50<Yaş≤ 60	4
60<Yaş≤ 70	5
70<Yaş≤ 85	6



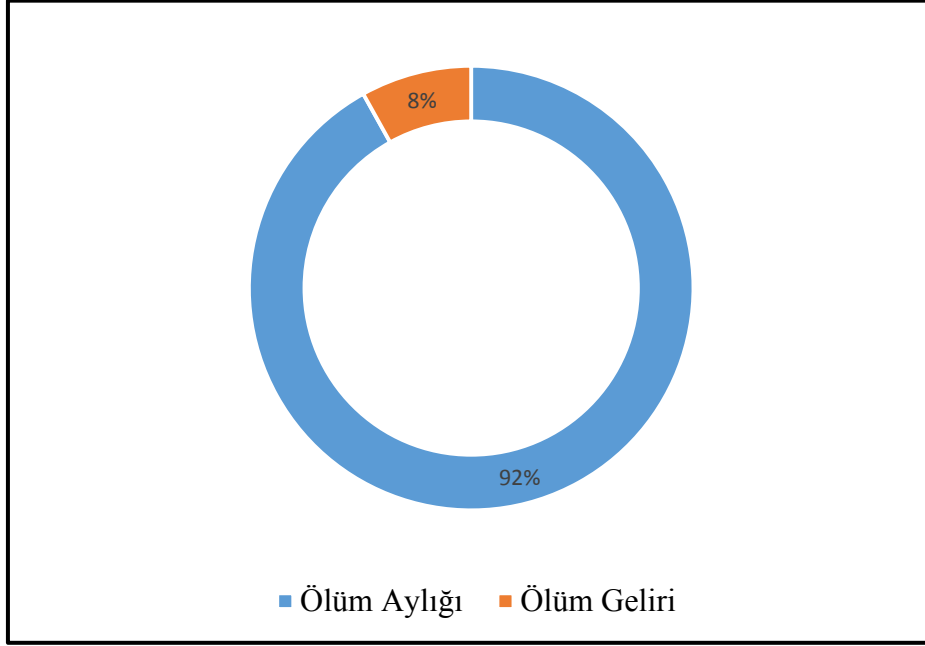
Şekil 5.1. Yaş aralıklarına göre kişilerin dağılımı

Aylık TÜRÜ: Bağlanan aylığın türünü ifade etmektedir. Ölüm geliri iş kazası sonucunda ölenlerin hak sahiplerine bağlanan ya da iş kazası sonucunda sakatlık derecelerine göre sürekli iş göremezlik ödemesi alan kişilerin ölümü sonucunda hak sahiplerine yapılan sürekli ödemeyi ifade etmektedir. Ölüm aylığı ise ölüm sigortasından hak sahiplerine sağlanan sürekli ödemeyi ifade etmektedir.³

³ 5510 sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu

Çizelge 5.3. Aylık türü değişkenine verilen kodlar

Aylık Türü	Kategoriler
Ölüm Aylığı	1
Ölüm Geliri	2



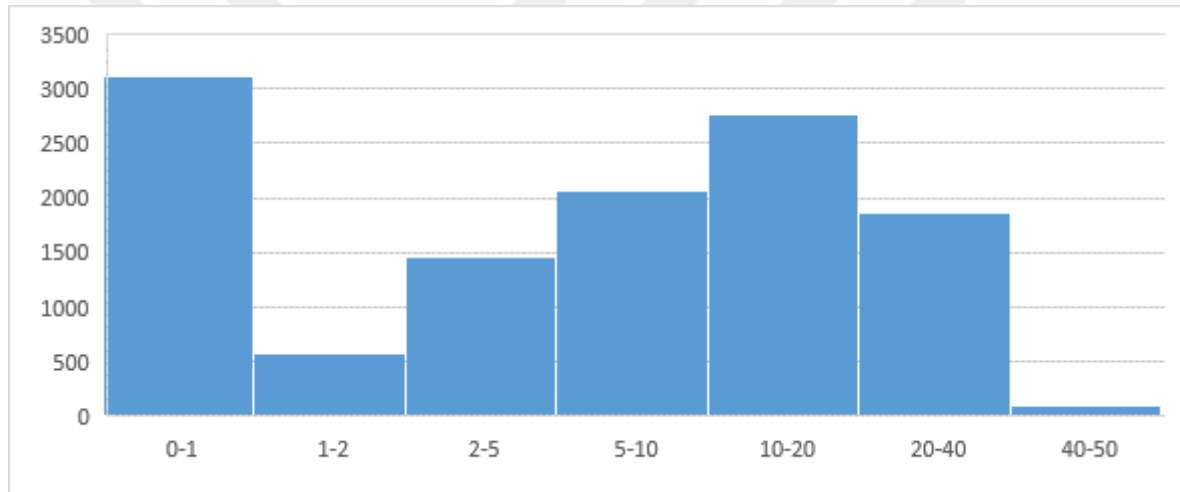
Şekil 5.2. Aylık türü değişkenine göre kişilerin yüzdesel dağılımı

Anlaşmalı boşandığı tespit edilen kişilere %92 ile çok büyük bir oranda ölüm aylığı bağlandığı, %8'ine ise ölüm geliri bağlandığı anlaşılmaktadır.

Ölüm-Boşanma Süresi: Anlaşmalı boşanan kişiye bağlanan aylığın sahibi olan sigortalının (ana veya baba) ölüm tarihi ile anlaşmalı boşanan kişiye bağlanan aylığın başlangıç tarihi arasında geçen süreyi ifade etmektedir. Baba veya ana öldükten ne kadar süre sonra yetim kızın boşandığını göstermektedir. Böyle bir süre Sosyal Güvenlik Kurumu veri tabanında tutulmamakta olup bu araştırma için hesaplanarak elde edilmiştir. Elde edilen bu yeni değişken ölçüm değeri yıl şeklinde ifade edilen sayısal değerler almaktadır. Uygulamada kullanılmak üzere 7 kategoriye ayrılmıştır.

Çizelge 5.4. Ölüm-boşanma süresinin kategorik veriye dönüştürülmesi

Ölüm-Boşanma Süresi (Yıl)	Kategoriler
$0 < \text{Ölüm ile Boşanma Arasında Geçen Süre} \leq 1$	1
$1 < \text{Ölüm ile Boşanma Arasında Geçen Süre} \leq 2$	2
$2 < \text{Ölüm ile Boşanma Arasında Geçen Süre} \leq 5$	3
$5 < \text{Ölüm ile Boşanma Arasında Geçen Süre} \leq 10$	4
$10 < \text{Ölüm ile Boşanma Arasında Geçen Süre} \leq 20$	5
$20 < \text{Ölüm ile Boşanma Arasında Geçen Süre} \leq 40$	6
$40 < \text{Ölüm ile Boşanma Arasında Geçen Süre} \leq 50$	7



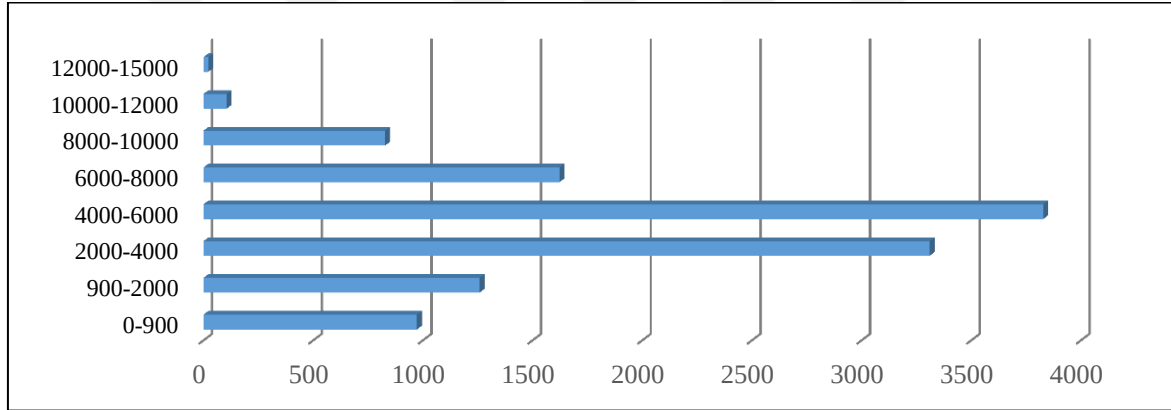
Şekil 5.3. Ölüm-Boşanma süresine ilişkin histogram

Şekil 5.3'e bakıldığında tüm anlaşmalı boşanmaların %26'sının ilk bir yılda yapıldığı %43'ünün de ilk 5 yıl içerisinde yapıldığı anlaşılmakla beraber bu davranışın büyük oranda ebeveynlerin ölüm tarihinden kısa bir süre sonra yapıldığı anlaşılmaktadır. 1-2 yıl arasında gerçekleşen boşanmaların sayısının ise diğer kategorilere göre %4,88 ile oldukça düşük olduğu görülmektedir.

Prim Ödeme Gün Sayısı: Kendisinden aylık bağlanan sigortalının (ana veya baba) prim ödeme gün sayısını ifade etmektedir. Emekli aylıklarının hesabında prim ödeme gün sayısının hem aylığa hak kazanma açısından hem de hesaplanacak aylığın miktarını belirlemesi açısından önemi büyüktür. Niceliksel bir değişken olan prim ödeme gün sayısı kategorik veri haline dönüştürülmüştür. Bu doğrultuda uygulamada kullanılmak üzere bağlanan aylıklara ait prim ödeme gün sayıları 8 kategoride ifade edilmiştir.

Çizelge 5.5. Prim ödeme gün sayısının kategorik veriye dönüştürülmesi

Prim Aralıkları(gün)	Kategoriler
0 < Prim Ödeme Gün Sayısı $\leq$ 900	1
900 < Prim Ödeme Gün Sayısı $\leq$ 2000	2
2000 < Prim Ödeme Gün Sayısı $\leq$ 4000	3
4000 < Prim Ödeme Gün Sayısı $\leq$ 6000	4
6000 < Prim Ödeme Gün Sayısı $\leq$ 8000	5
8000 < Prim Ödeme Gün Sayısı $\leq$ 10000	6
10000 < Prim Ödeme Gün Sayısı $\leq$ 12000	7
12000 < Prim Ödeme Gün Sayısı $\leq$ 15000	8



Şekil 5.4. Prim ödeme gün sayısına ilişkin frekansların dağılımı

Şekil 5.4.'e göre anlaşmalı boşanma kapsamında kesilen aylıkların dağılımı gösterilmektedir. En fazla yoğunluğun 2000-4000 gün arasında ve 4000-6000 gün arasında olduğu anlaşılmaktadır.

Banka: Anlaşmalı boşanan sigortalının aylık ya da gelir aldığı bankayı ifade etmektedir. Eşinden anlaşarak aylık alırken tespit edilen kadınların ödemelerini 16 ayrı bankadan aldıkları anlaşılmıştır. Analizde kullanmak üzere bankalara 1 ile 16 arasında kodlar verilmiştir.

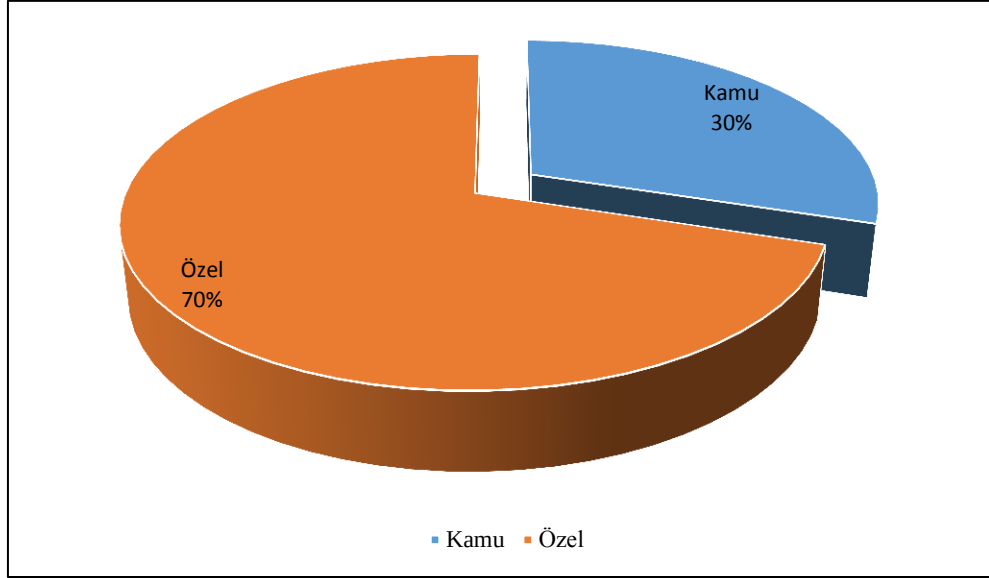
Çizelge 5.6. Bankalara verilen kodlar

Banka	Kategoriler
Akbank	1
Burganbank	2
Denizbank	3
Finansbank	4
Garanti Bankası	5
Halkbank	6
HSBC	7
INGBank	8
İş Bankası	9
Kuveyt Türk	10
PTT	11
TEB	12
Şekerbank	13
Vakıflar Bankası	14
Yapı kredi Bank	15
Ziraat Bankası	16

Sektör: Kendisinden aylık bağlanan kişinin emekli olduğu sektörü ifade etmektedir. Sektör değişkeni kategorik bir değişken olup, programda kamu sektörü için 1 özel sektör için 2 kodu kullanılmıştır.

Çizelge 5.7. Sektör sınıflaması

Sektör	Kategoriler
Kamu	1
Özel	2



Şekil 5.5. Bağlanan aylıkların sektöre göre dağılımı

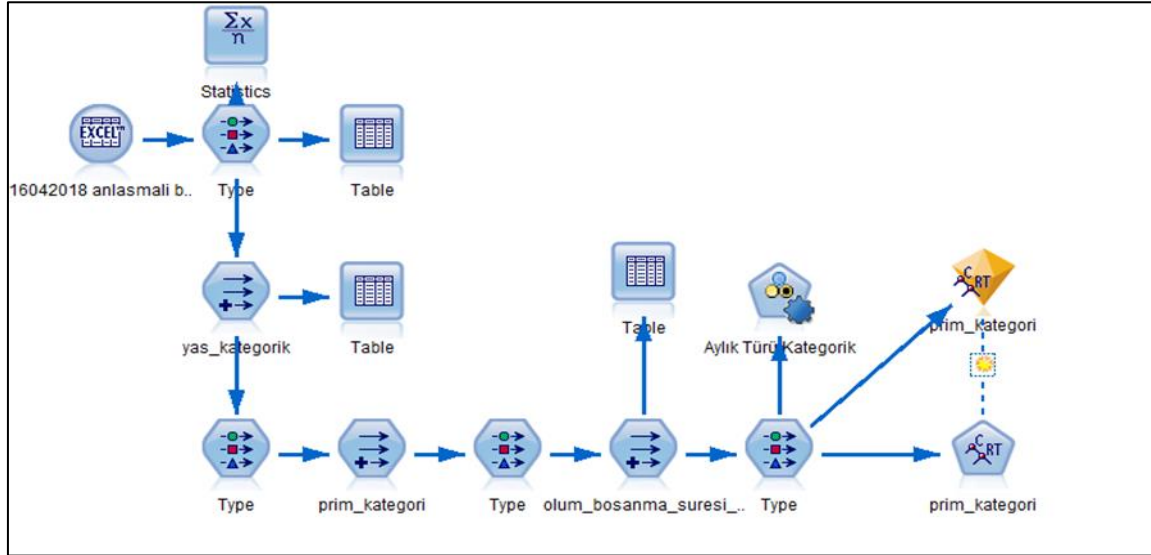
Sektör bazında frekanslara bakıldığında araştırmaya konu olan aylık veya gelirlerin %70'inin özel sektörden %30'unun ise kamu sektöründen bağlanan aylıklar olduğu görülmektedir.

Açıklanan bu değişkenlere veri madenciliği tekniklerinden karar ağacı algoritmalarından CART algoritması uygulanarak anlaşmalı boşananlar arasındaki ortak özellikler ortaya çıkarılmıştır. Bu sayede olup henüz tespit edilemeyen kesim için bir model elde edilmiştir. Bu sayede Sosyal Güvenlik Kurumu için anlaşmalı boşanan kız çocuklarının davranış modeli belirlenmiştir. Elde edilen sonuçlar Kurumun denetimle görevlendirilmiş memurlarının bu kişileri tespit edebilmesini kolaylaştıracak bir sonuç ortaya konulması açısından önemlidir.

5.2. Uygulamada İzlenen Adımlar

SPSS Modeler 18. Programı açıldıktan sonra ilk etapta boş stream ekranı bulunmaktadır. Analize başlarken ilk olarak kayıtların yer aldığı excel modeller programına okutulmuştur. Sonrasında yaş, prim ödeme gün sayısı ve ölüm-boşanma süresi sürekli değişkenleri type düğümü yardımıyla kategorik veriye dönüştürülmüştür. Daha sonra sınıflama algoritmalarından CART algoritmasını ifade eden CRT düğümü akışa bağlanmış ve burada prim ödeme gün sayısı bağımlı değişken, sektör, aylık türü, ölüm-boşanma süresi, il, banka

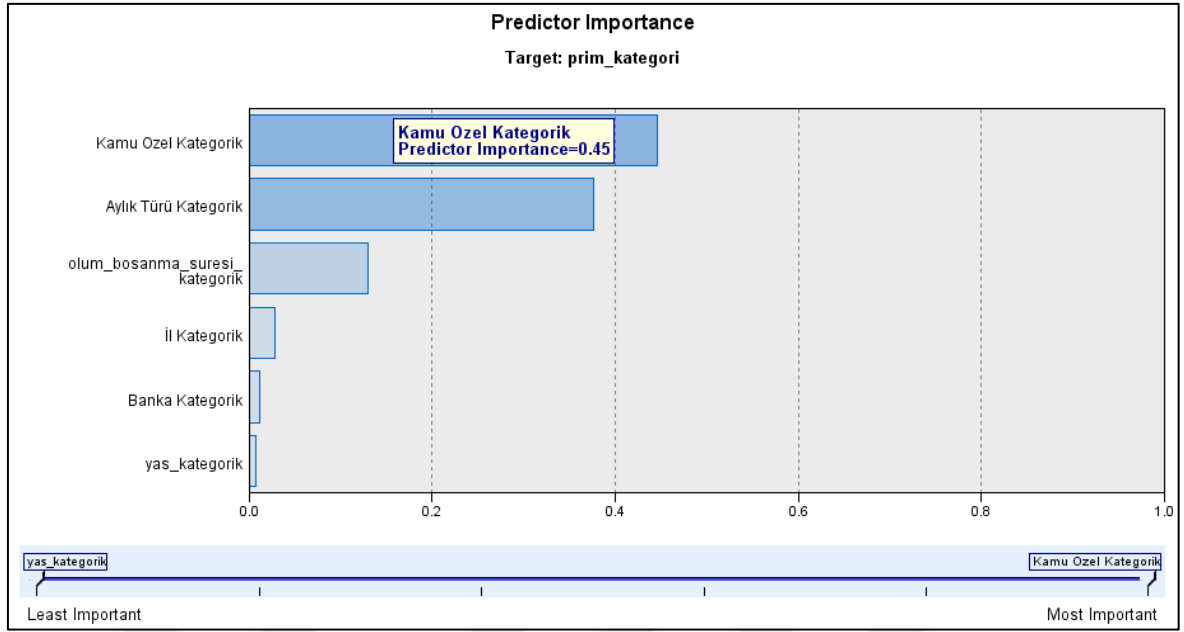
ve yaş değişkenleri de bağımsız değişken olarak seçildikten sonra stream tamamlanarak analiz sonuçlandırılmıştır. Programda kullanılan akış Şekil 5.6'da gösterilmiştir.



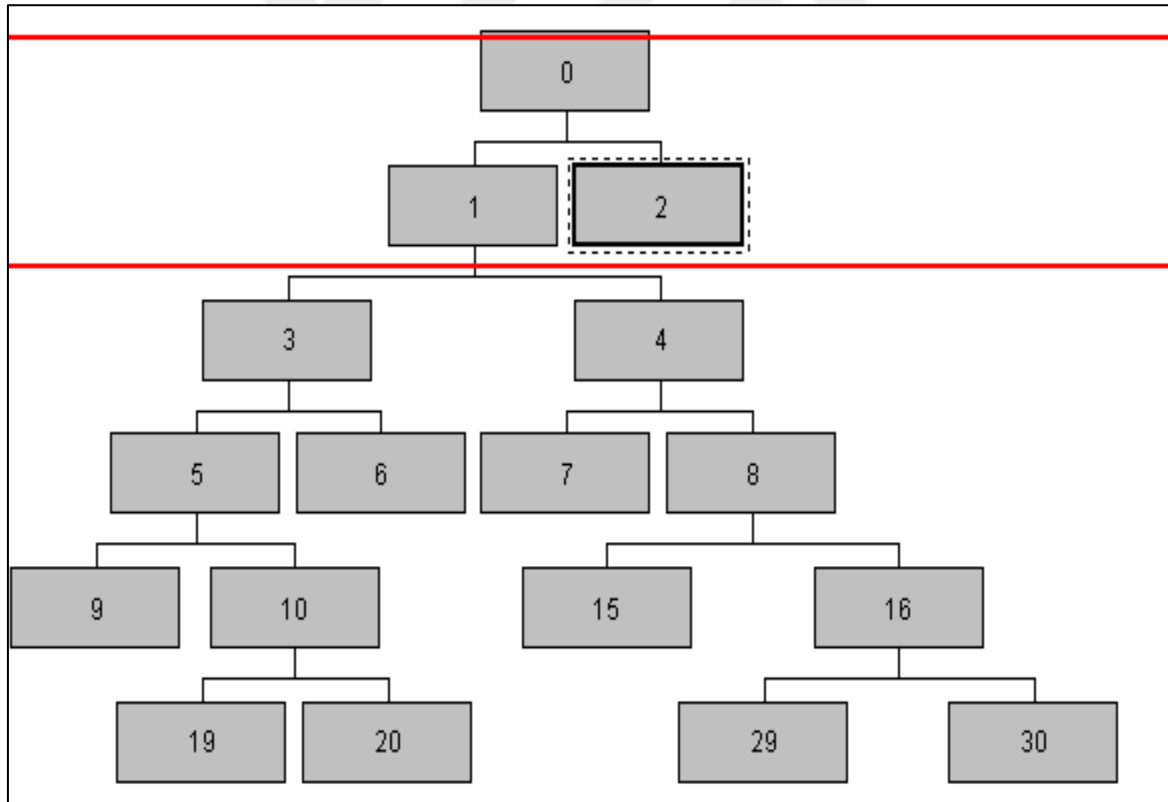
Şekil 5.6. SPSS modeler programında yapılan analizin akış şeması (stream)

5.3. Karar Ağacının Elde Edilmesi

CART algoritması kullanılarak kurulan modelde prim ödeme gün sayısı değişkeni bağımlı değişken olarak, sektör, aylık türü, ölüm-boşanma süresi, il, banka ve yaş değişkenleri ise modelde açıklayıcı değişken olarak seçilmiştir. Şekil 5.7'den, açıklanan değişken prim ödeme gün sayısı üzerinde en önemli değişkenlerin sırasıyla sektör, aylık türü, ölüm-boşanma süresi, il, banka ve yaş değişkenleri olduğu görülmektedir. Bu değişkenlere göre elde edilen karar ağacının genel görünümü Şekil 5.8'de gösterilmektedir.



Şekil 5.7. CART algoritmasına göre prim değişkeni üzerindeki önemli değişkenler



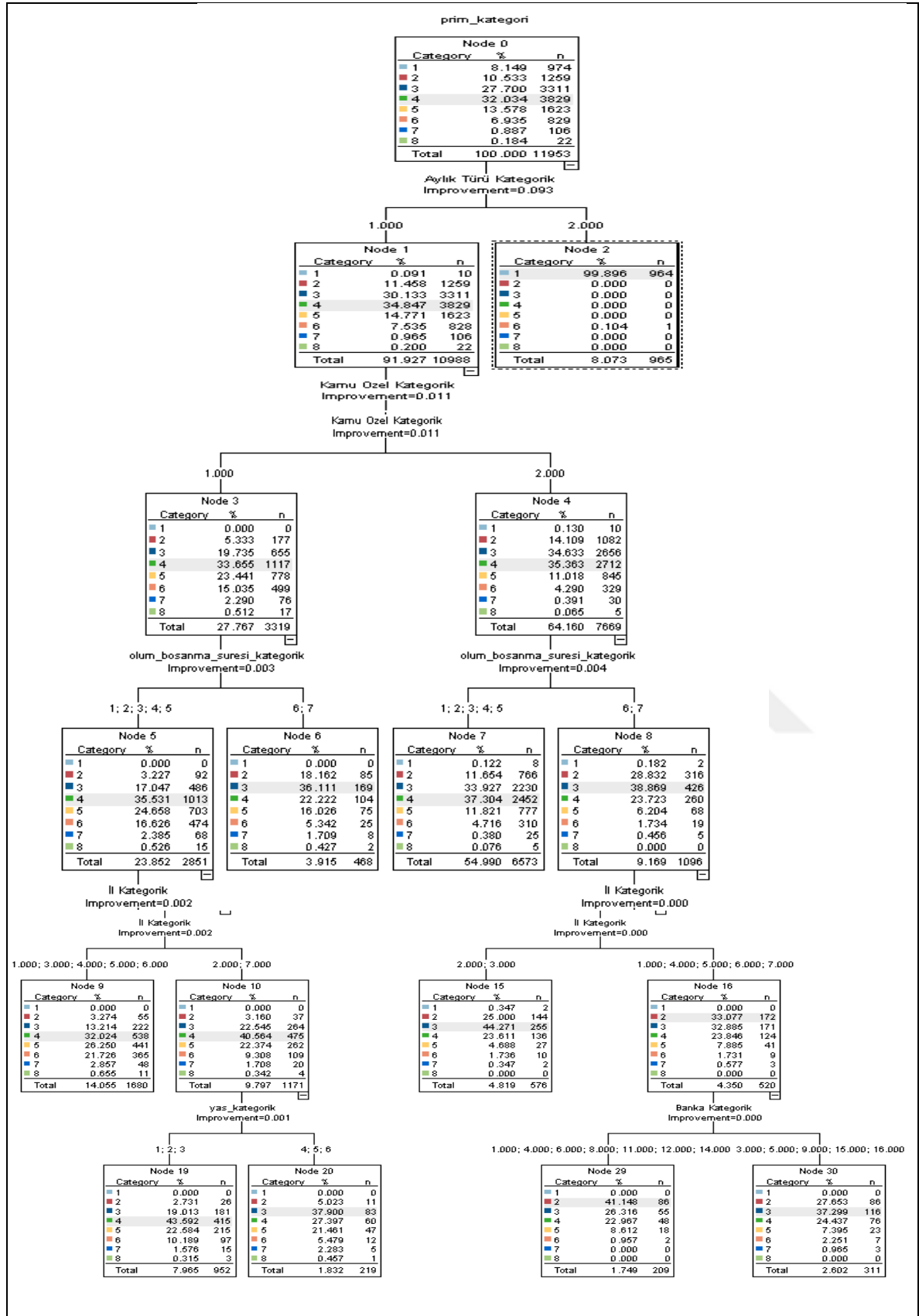
Şekil 5.8. CART algoritması ile elde edilen karar ağacının genel görünümü

Karar ağacı, ilk iki düğümün yer aldığı bölüme “A”, 3 ve 8 arasında yer alan düğümlerin bulunduğu bölüme “B” ve sonraki düğümlerin yer aldığı bölüme de “C” adı verilerek açıklanacaktır. Ayrıca karar ağacından anlaşılmalı boşandığı tespit edilen kişilere ait kök

düğümünden yaprak düğüme 9 davranış profili oluştuğu görülmektedir. Bu profiller çizelge 5.8.'de özetlenmiştir. Çizelge 5,8'de profil 8, profil 9 ve profil 3 sınıfladığı kişi sayısının oranına göre öne çıkmaktadır.

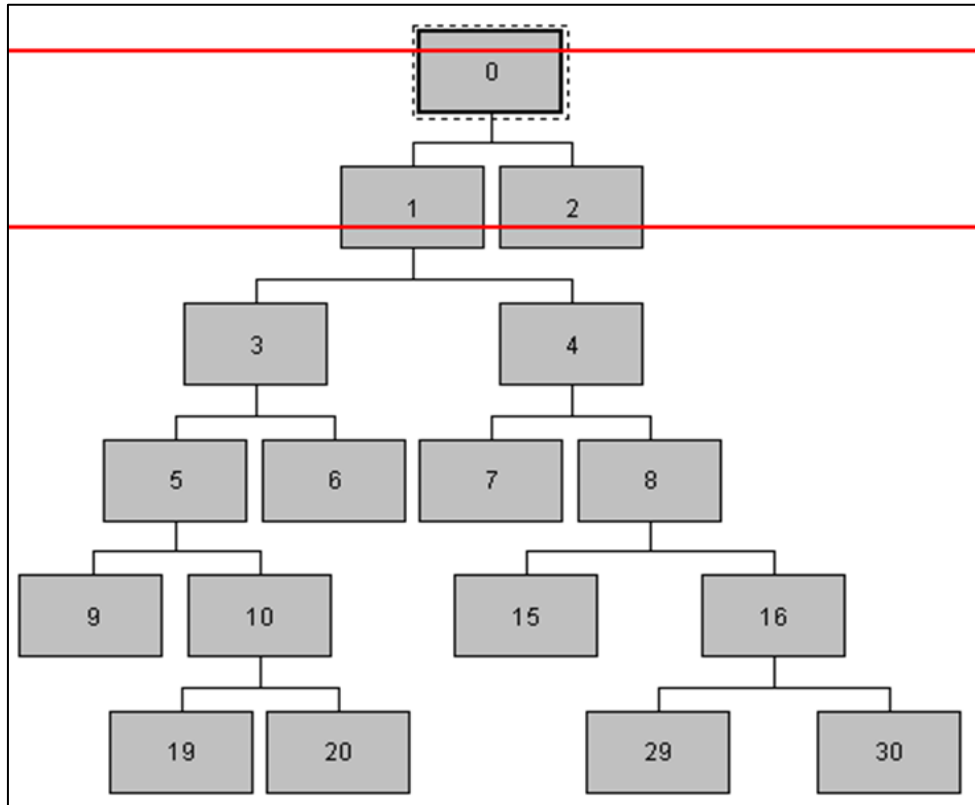
Çizelge 5.8. Sınıflama sonucunda elde edilen karar ağacında oluşan davranış profilleri

Davranış Profilleri	Düğüm/ Yapraklar	Aylık Türü	Sektör	Ölüm Boşanma Süresi	Bölge	Yaş	Banka
Davranış Profili 1	Node 19	1	1	1,2,3,4,5	2,7	1,2,3	
Davranış Profili 2	Node 20	1	1	1,2,3,4,5	2,7	4,5,6	
Davranış Profili 3	Node 9	1	1	1,2,3,4,5	1,3,4,5,6		
Davranış Profili 4	Node6	1	1	6,7			
Davranış Profili 5	Node 29	1	2	6,7	1,4,5,6,7		1,4,6,8,11,12,14
Davranış Profili 6	Node 30	1	2	6,7	1,4,5,6,7		3,5,9,15,16
Davranış Profili 7	Node 15	1	2	6,7	2,3		
Davranış Profili 8	Node 7	1	2	1,2,3,4,5			
Davranış Profili 9	Node 2	2					

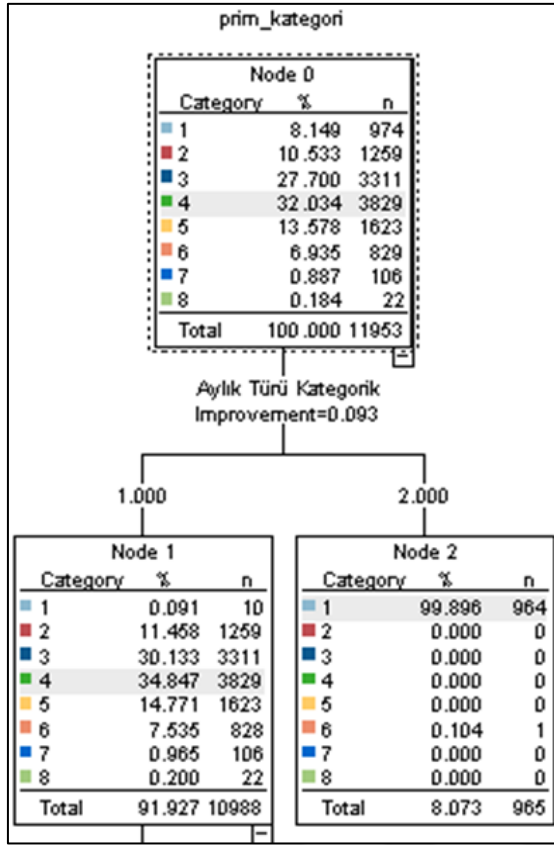


Şekil 5.9. Anlaşmalı boşananlara ilişkin karar ağacının genel görünümü

Şekil 5.9.'da gösterilen CART algoritmasına göre oluşan karar ağacının kök düğümünden anlaşmalı boşandığı tespit edilen kadınların %8,149'una 0-900 gün arasında, %10,533'üne 900-2000 gün arasında, %27,7'sine 2000-4000 gün arasında, %32,034'üne 4000-6000 gün arasında, %13,578'ine 6000-8000 gün arasında ve geri kalan %8'ine ise 8000-15000 prim ödeme gün sayısı üzerinden aylık bağladığı görülmektedir. Kök düğümüne bakıldığında en büyük iki grupta anlaşmalı boşanan kadınlara bağlanan aylıkların %27,7 orana sahip 2000-4000 gün ile %32,034 orana sahip 4000-6000 gün arasında olduğu görülmektedir. Karar ağacının A, B ve C bölümleri şekil 5.11. şekil 5.13. ve şekil 5.15.'te verilerek açıklanmıştır.

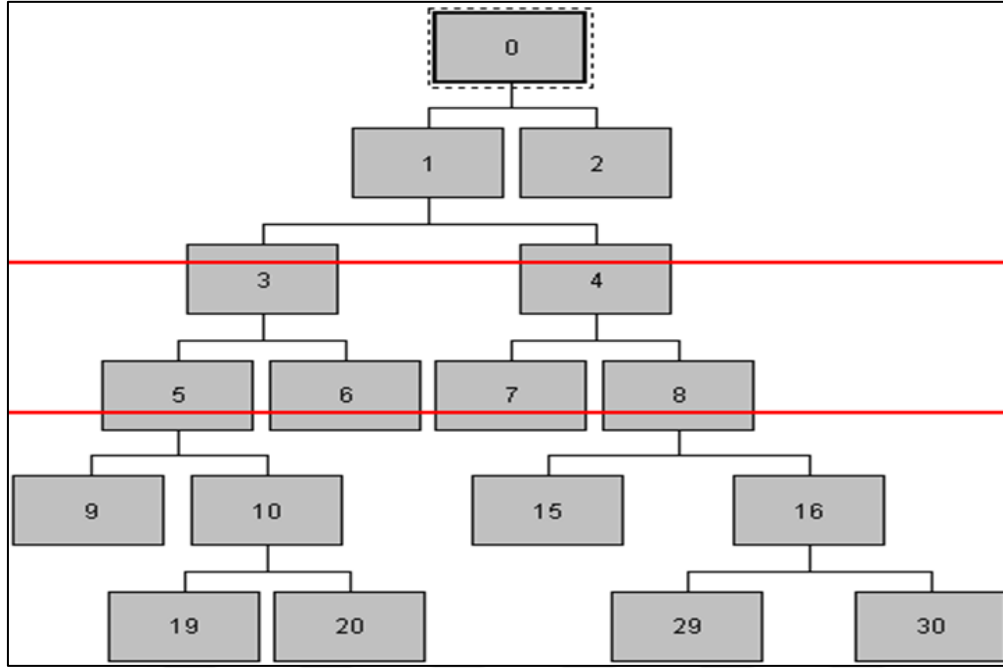


Şekil 5.10. Karar ağacı A bölümü

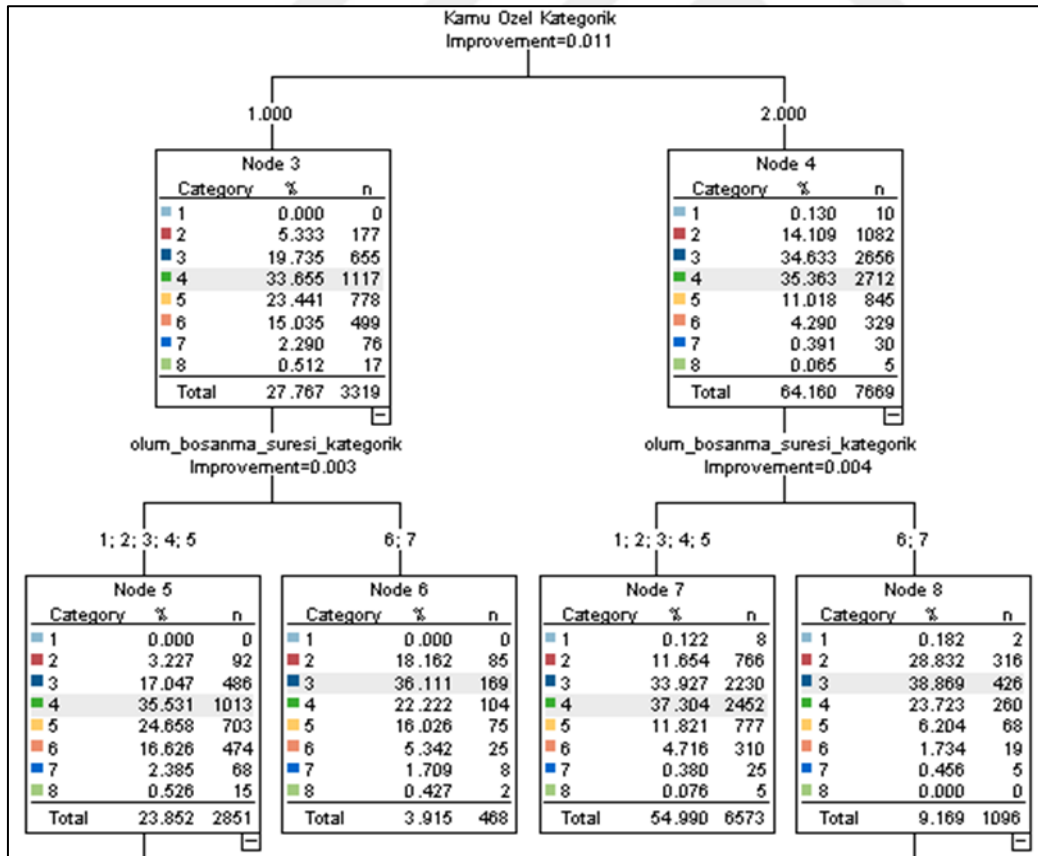


Şekil 5.11. Karar ağacı A bölümünü oluşturan düğümler

Kök düğümünden sonra ilk sınıflamanın bağlanan aylıkların türüne göre olduğu görülmektedir. Anlaşmalı boşanan kadınlara bağlanan aylıkların %91,927'sinin "Aylık" türünde, %8,073'ünün ise "Gelir" türünde olduğu anlaşılmaktadır. Yine bu kapsamda bağlanan aylıklardan sırasıyla 2000-4000 gün arasında olanların oranı %30,133 olarak, 4000-6000 gün arasında olanların oranı ise %34,847 olarak hesaplanmıştır. 9. uncu davranış profilini oluşturan ikinci düğümde "Gelir" bağlananların neredeyse %100'ünün birinci kategoride 0-900 arasında prim ödeme gün sayısında sınıflandığı görülmektedir. İkinci düğüm için ayrıca, çalışma hayatında herkesin çalışma hayatı süresince iş kazası geçirebileceği göz önüne alındığında "Gelir" aldıktan sonra anlaşmalı boşanma yapan kız çocuklarının kendilerine "Gelir" bağlanan anne veya babalarının çalışma hayatlarının ilk yıllarında geçirdikleri iş kazasına bağlı olarak "Gelir" aldıkları anlaşılmaktadır.



Şekil 5.12. Karar ağacı B bölümü

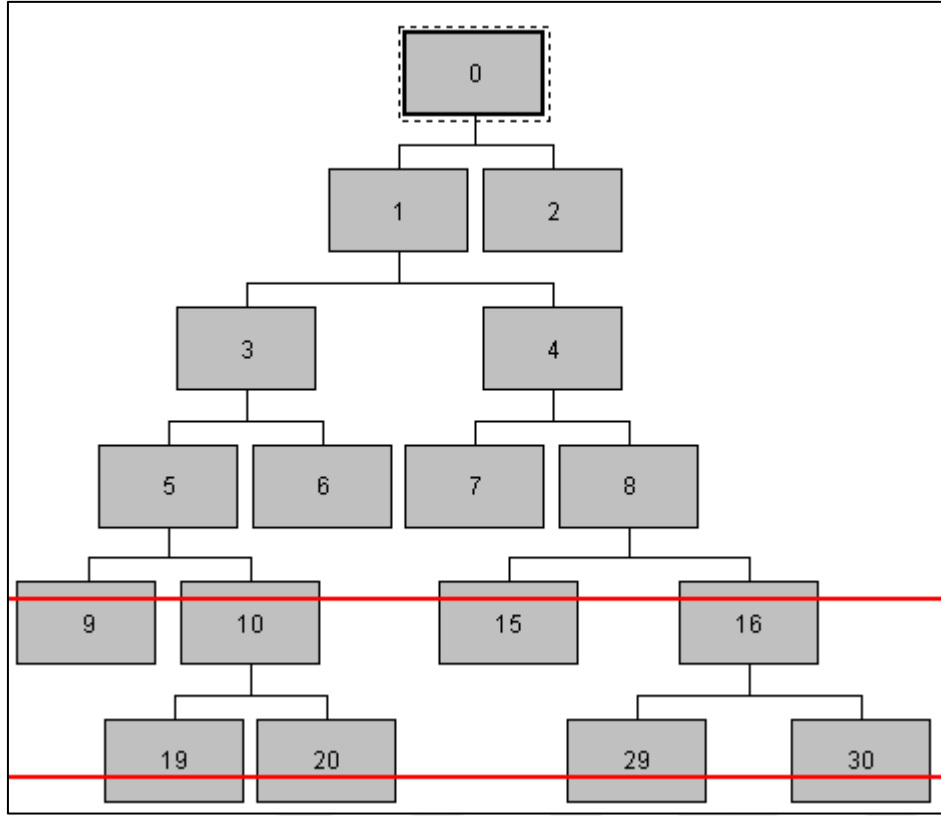


Şekil 5.13. Karar ağacı B bölümünü oluşturan düğümler

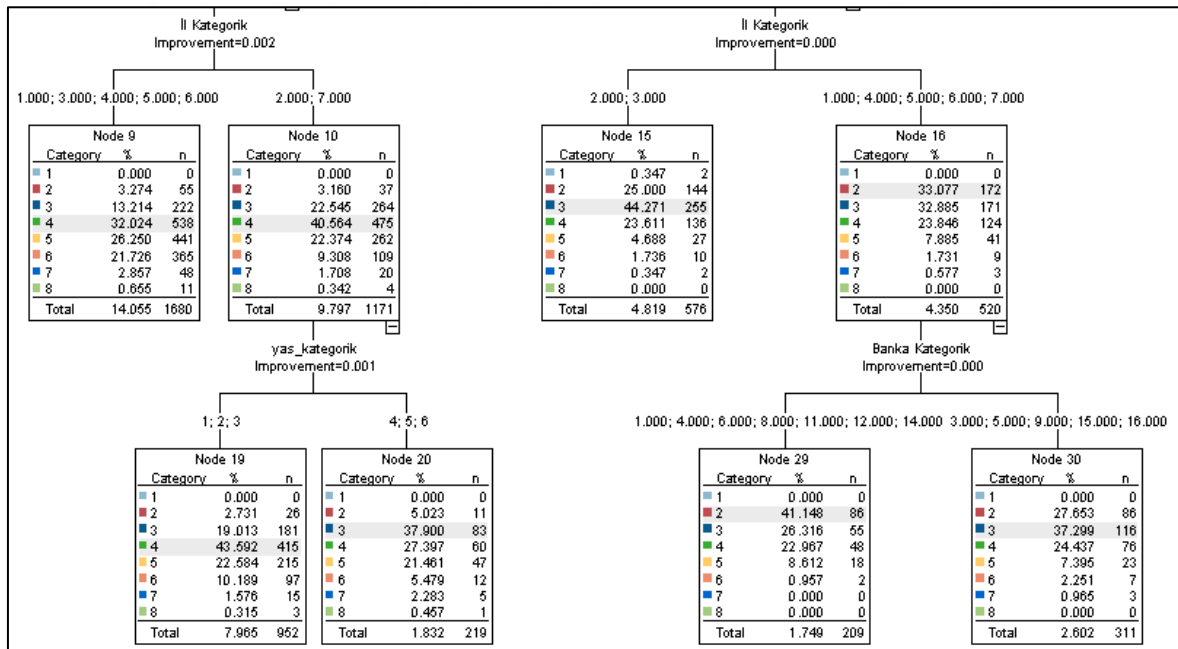
Şekil 5.13'den toplam içerisinde %92 orana sahip aylık türündeki anlaşmalı boşanmalar sektöre göre ayrıldığında bunlardan Kamu sektöründen aylık bağlananlar %27,767, Özel sektörden aylık bağlananların ise %64,160 olarak hesaplanmıştır. Burada sektör fark etmeksizin prim ödeme gün sayısı açısından 3 üncü ve 4 üncü kategorilerde yoğunlaşma gözlemlenmiştir. Kamu sektöründen bağlanan aylıklarda Prim ödeme gün sayısı 4000-6000 gün arasında olanlar %33,655, özel sektörden bağlanan aylıklarda ise yine %35,363 ile tüm kategoriler içerisinde en yüksek oranda aylıkların 4000-6000 gün arasından bağlandığı gözlemlenmektedir.

Kamu sektöründen bağlanan aylıklara sahip kız çocuklarının anlaşmalı boşanma davranışını anne veya baba öldükten sonra ilk 20 yılda gösterenlerin oranı %23,8, 20 yıl ve sonrasında gösterenlerin oranı ise %3,9 şeklindedir. 0-900 gün arasında prim ödeme gün sayısına sahip aylıklardan (Vazife malullüğü aylıkları gibi belirli prim ödeme gün sayısı gerektirmeyen) anlaşmalı boşanma olgusuna hiç rastlanılmamıştır. Yine kamu sektörü için anlaşmalı boşananlara bağlanan aylıkların 4000 gün ve üzerinde yüksek prim ödeme gün sayılarında gerçekleştiği görülmektedir.

Özel sektörden bağlanan aylıklara sahip kız çocuklarının ise anlaşmalı boşanma davranışını anne veya baba öldükten sonra ilk 20 yılda gösterenlerin oranı %54,99, 20 yıl ve sonrasında gösterenlerin oranı ise %9,16 şeklindedir. Buna ek olarak ölüm-boşanma sürelerinin frekanslarına bakıldığında ise tüm anlaşmalı boşanmaların %26'sının ilk bir yılda yapıldığı %43'ünün de ilk 5 yıl içerisinde yapıldığı gözlemlenmekle beraber anlaşmalı boşanma olgusunun büyük oranda ebeveynlerin ölüm tarihinden kısa bir süre sonra yapıldığı anlaşılmaktadır. Sonraki dallanmalar kamu sektöründen gelen ve ilk 20 yıl içerisinde boşanma gerçekleştiren %23,85'lik grup ile özel sektörden gelen ve 20-50 yıl süresi arasında boşanma yapan %9,16'lık grup üzerine gerçekleşmiştir.



Şekil 5.14. Karar ağacı C bölümü



Şekil 5.15. Karar ağacı C bölümünü oluşturan düğümler

Şekil 5.15'den, kamu sektöründen gelen ve ilk 20 yıl içerisinde boşanma gerçekleştiren %23,85'lik grubun bölgelere göre dağılımına bakıldığında %9,79 oranıyla Karadeniz ve

Doğu Anadolu bölgelerinin bir düğümde, %14,05'inin ise İç Anadolu, Marmara, Ege, Akdeniz ve Güney Doğu Anadolu bölgelerinin ise diğer düğümde sınıflandığı görülmektedir. Özel sektörden gelen %9,16'lık grubun %4,81 ile Karadeniz ve Marmara bölgelerinde, %4,35 ile İç Anadolu, Ege, Akdeniz, Güney Doğu Anadolu ve Doğu Anadolu Bölgelerinde sınıflanmıştır. Yine bu sınıflarda prim ödeme gün sayıları incelendiğinde, kamu sektöründen gelenlerin yüksek prim ödeme gün sayılarında anlaşmalı boşanma davranışını gösterdiği, özel sektörden gelenlere bakıldığında ise sırasıyla 900-2000 gün, 2000-4000 gün gibi daha düşük prim ödeme gün sayılarında %30 ila %44 gibi yüksek oranlarda sınıflandığı görülmektedir. Sonraki dallanmaların, bağımlı değişken prim ödeme gün sayısı üzerinde önem sırasına göre en sonda yer alan yaş ve banka değişkenlerinden gerçekleştiği görülmektedir. Kamu sektöründen gelen ve Karadeniz ve Doğu Anadolu bölgelerinde anlaşmalı boşanma yapan %9,79 oranındaki yetimin kız çocuğundan %7,9'unun 18-50 yaş aralığında olduğu ve %1,8'inin ise 50 yaş üzerinde sınıflandığı görülmüştür. Kamu sektöründen gelen bir dallanma olduğundan yine yüksek prim ödeme gün sayısına sahip aylıklardan oluştuğu sonucuna ulaşılmıştır.

6. SONUÇ VE ÖNERİLER

Son yıllarda gelişen teknolojiyle beraber eskiden kâğıt ortamında yapılan birçok iş ve işlem elektronik ortamda yürütülmeye başlanmış ve bunun sonucunda da şirketlerde ve kamu kurumlarında kişilere ait çok çeşitli yapıda ve çok büyük sayıda veriler tutulmaktadır. Kurumsal yapılarda çalışan yöneticiler ve karar alıcılar şirketleri ya da kurumlarının faaliyetlerinin ne yöne gittiğini tespit edebilmek ya da kendileri için risk oluşturabilecek nedenleri bilmek istemektedir. Bu sebeple işletme ya da kurumların kayıtlarında yer alan gizli bilgiyi ortaya çıkarmak için veri madenciliği uygulamaları kullanılmaktadır.

Bu araştırmanın konusunu oluşturan anne veya babasından aylık alabilmek için eşiyle anlaşmalı boşanıp, aynı evde yaşayanlar Sosyal Güvenlik Kurumu için risk teşkil eden bir unsur olarak kabul edilmektedir. Çünkü sosyal güvenlik kanunumuzda açıkça ödenmemesi gerektiği belirtilen bir paranın bu kapsamdaki kişilere ölene kadar ödenmesi, aktüeryal dengeleri olumsuz etkileyeceği açıktır. Anlaşmalı boşanma olgusunu gerçekleştiren kişilerin ortak özelliklerini bilmeyen karar alıcılar anlaşmalı boşanma olgusunun tespiti için daha fazla insan kaynağı ile daha gelenekçi ve bürokratik usullerle soruna çözüm aramaya çalışacaktır. Nitekim hali hazırda uygulama bu yöndedir. Bu durum ise çok fazla iş gücü kaybına ve zaman kaybına neden olacağından işletmeler ve kurumlar için bir başka risk faktörü olarak değerlendirilebilir.

Anlaşmalı boşanarak eşiyle aynı evde yaşayan kişilerin tespiti zor bir uygulamadır. Sosyal Güvenlik Kurumu'nun denetim ve kontrolle görevli memurlarının yani denetmenlerin karar alma süreçlerini kolaylaştırmak amacıyla daha önce yapılan denetimler sonucunda tespit edilen kişilerin Sosyal Güvenlik Kurumu veri tabanında tutulan karakteristik özelliklerinden yararlanılarak bahsedilen probleme çözüm olması açısından bu gruba ilişkin veriler içinde saklanan gizli kalmış örüntünün ortaya çıkarılması Kurum açısından oldukça önemlidir.

Bu amaç doğrultusunda, bu çalışmada veri madenciliği sınıflama tekniklerinden veri yapısına uygun olması ve uygulamalarda en sık kullanılan yöntemlerden biri olması nedenleriyle veri madenciliği sınıflama tekniklerinden CART algoritması kullanılarak bir karar ağacı oluşturulmuştur. Böylece aylıkları anlaşmalı boşanma kapsamında kesilen

kişilerin Sosyal Güvenlik Kurumu veri tabanlarında tutulan karakteristik özelliklerinin sınıflandırılmasıyla bu kişilerin davranış modeli ortaya konulmuştur.

Anlaşmalı boşanma yapan kişilerle alakalı önceden bilinmeyi, akla gelmeyi, tahmin edilemeyi ortaya koymak ve bu kişilere ait veriler üzerine yapılan analiz anlamlı ve kullanışlı bir şekilde özetlenmiştir. Bu çalışma sonucunda elde edilen bulgular aşağıdaki şekilde sıralanabilmektedir.

Anlaşmalı Boşanmaların % 91,9'unun ölüm aylıklarından oluştuğu, %8inin ise ölüm gelirinden oluştuğu tespit edilmiştir.

Tüm düşümler dikkate alındığında anlaşmalı boşanma davranışı gösterenlere % 30 oranında 2000-4000 gün üzerinden, %30-%40 bandında değişen oranlarla da 4000-6000 gün üzerinden aylıklar hesaplandığı gözlemlenmiştir.

%92 orana sahip aylık türündeki anlaşmalı boşanmalar sektöre göre ayrıldığında bunlardan Kamu sektöründen bağlananlar %27,7, Özel sektörden bağlananların ise %64,1 olarak hesaplanmıştır. Buradan anlaşmalı boşanma gösteren kişilerin büyük oranda özel sektörden emekli kişilerin çocukları olduğu anlaşılmaktadır.

Kamu sektöründen bağlanan aylıklara sahip kız çocuklarının anlaşmalı boşanma davranışını anne veya baba öldükten sonra ilk 20 yılda gösterenlerin oranı %23,8, 20 yıl ve sonrasında gösterenlerin oranı ise %3,9 şeklindedir. Özel sektörden bağlanan aylıklara sahip kız çocuklarının ise anlaşmalı boşanma davranışını ana veya baba öldükten sonra ilk 20 yılda gösterenlerin oranı %54,99, 20 yıl ve sonrasında gösterenlerin oranı ise %9,16 şeklinde hesaplanmıştır. Ölüm-boşanma sürelerinin frekanslarına da bakıldığında tüm anlaşmalı boşanmaların %26sının ilk bir yılda yapıldığı %43'ünün de ilk 5 yıl içerisinde yapıldığı anlaşılmakla beraber bu davranışın büyük oranda ebeveynlerin ölüm tarihinden kısa bir süre sonra yapıldığı anlaşılmaktadır.

Anlaşmalı boşanmaların tespitinin birçok yazışma, yerinde denetim ve ifadelere başvurma gibi bürokratik süreçlere sahip olduğu, buna rağmen tespitinin zorluğu uzun sürdüğü göz önüne alındığında; Daha önce tespit edilen grubun davranış modeli araştırma sonucunda elde edilmiştir. Araştırma sonucunda elde edilen bulgular karar alma süreçlerini

kolaylaştıracak niteliktedir. Örneğin Bir denetmen denetim yaptığı kişi ile alakalı kendilerine bakılması talimatlandırılan hususları incelediğinde ortada anlaşmalı boşanma durumunu gösteren somut bir kanıt bulamadığında ve bu kişilerle alakalı çelişkili ifadeleri nedeniyle şüphelendiği düşünülürse, araştırma konusundan haberdar bir denetmen şikayete konu kişinin özelliklerini inceleyerek yapılan yardımın ölüm aylığı, prim ödeme gün sayısının 2000-3000 gün arasında olduğunu, özel sektör aylığı olduğunu ve boşanmanın da ölümden sonra ilk 1 yıl ya da ilk 5 yıl içerisinde gerçekleştiğini gördüğünde bu kişilerin anlaşmalı boşanma yapıp yapmadığının tespitinde daha kolay karar verebilecektir. Bu tür özelliklere sahip şüpheli kişiler hakkında daha fazla araştırma yapmaları gerektiği konusunda denetim personellerine ön bilgi sunulmuştur.

Bu çalışma ile Sosyal Güvenlik Kurumu'ndan aylık alırken boşanan ve anlaşmalı boşanma yapıp yapmadığı bilinmeyen (ihbar ve şikâyet olmadığı ya da rutin denetimlere takılmayan) yetim kız çocukları hakkında bir model elde edilmiştir. Bu konu ile ilgili gelecekte bilinmeyi tespit etmek amacıyla Sosyal Güvenlik Kurumu veri tabanında tutulan kayıtlar üzerine bu araştırmada elde edilen model uygulandığında potansiyel risk grubunun haritası da çıkarılabilecektir.



KAYNAKLAR

- Agrawal, R., Imieliński T., Swami A. (1993). *Mining association rules between sets of items in large databases*. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 22(2), 207-216.
- Akça, F. (2014). *Veri Madenciliği İle Fen Fakülteleri Öğrenci Profillerinin İncelenmesi: Gazi Üniversitesi Örneği*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Alan, M.A. (2014). Karar ağaçlarıyla öğrenci verilerinin sınıflandırılması. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 28(4), 101-112.
- Aydın, S. (2007). *Veri Madenciliği Ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama*, Yüksek Lisans Tezi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Eskişehir.
- Ayık, Y., Z., Özdemir, A., Yavuz, U. (2007). Lise türü ve lise mezuniyet başarısının, kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441-454.
- Baradwaj, B. K., Pal, S. (2011). Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, 2(6), 63-69.
- Beyer K., Goldstein J., Ramakrishnan R., Shaft U., (1999). *When Is 'Nearest Neighbor' Meaningful?*, 7. International Database Theory Conference Proceedings, 217-235.
- Budak, E., Ç. (2013). Üniversite öğrencileri için bilgisayar okuryazarlığını etkileyen faktörlerin etkisinin veri madenciliği ile analizi. *Online Academic Journal of Information Technology*, 4(11), 58-70.
- Chen, M. S., Han, J., Yu, P. S. (1996). Data mining: an overview from a database perspective. *Transactions on Knowledge and Data Engineering*, 8(6), 866-883.
- Çakır, F. (2012). *Veri Madenciliği Yöntemleri Kullanılarak Küçük Ve Orta Ölçekli İşletmelere İlişkin Bazı Verilerin Çözümlemesi*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Çığışar, B. (2017). *Kredi Risklerinde Veri Madenciliği Sınıflama Algoritmaları*, Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.
- Değirmenci, T. (2014). *Resmi İstatistiklerde Veri Madenciliği Yaklaşımı*, Yüksek Lisans Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Kayseri.
- Demirel, B. (2010). *Veri Madenciliğinde CHAID Algoritmasının Sosyal Güvenlik Kurumu Veri Tabanına Uygulanması*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Dumitru, D. (2009). Prediction of recurrent events in breast cancer using the Naive Bayesian classification. *Annals of University of Craiova Mathematics and Computer Science Series*, 36(2), 92-96.

- Diler, S. (2016). *Veri Madenciliği Süreçleri Ve Karar Ağaçları Algoritmaları İle Bir Uygulama*, Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van.
- Du, W., Zhan, Z. (2003). *Using randomized response techniques for privacy-preserving data mining*. Proceedings of International Conference on Knowledge Discovery and Data Mining, 505-510.
- Dunham, M., H. (2003). *Data Mining Introductory and Advanced Topics*. New Jersey: Pearson Education, 2, 3, 5, 97, 98, 140, 142.
- Ekdemir, Ş., H. (2005). *Avrupa Birliği- Türkiye Sosyal Güvenlik Normları*. Ankara: Ajans-Türk Basın ve Basım, 105.
- Emel, G., G., Taşkın, Ç. (2005). Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 222-239.
- Fausett L. (1994). *Fundamentals of Neural Networks: Architectures, algorithms, and applications*. New Jersey: Prentice-Hall, 3.
- Gemici, B. (2012). *Veri Madenciliği Ve Bir Uygulaması*, Yüksek Lisans Tezi, T.C. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- Guha S., Rastogi R., Shim K. (1998), CURE: An efficient clustering algorithm for large databases, *Newsletter ACM SIGMOD Record*, 27(2), 73-84.
- İnternet: Fazla veya Yersiz Ödemelerin Tahsiline İlişkin Usul ve Esaslar Hakkında Yönetmelik. URL: <http://www.mevzuat.gov.tr/Metin.Aspx?MevzuatKod=7.5.12463&MevzuatIliski=0&sourceXmlSearch>, Son Erişim Tarihi: 10.05.2019.
- İnternet: 5510 Sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu. URL: <http://www.mevzuat.gov.tr/MevzuatMetin/1.5.5510.pdf>, Son Erişim Tarihi: 10.05.2019.
- İnternet: Yapay Sinir Ağı Resmi. URL: <https://www.kimnezamanicatetti.com/yapay-sinir-agi/>, Son Erişim Tarihi: 10.05.2019.
- İnternet: Sosyal Güvenlik Denetmenleri Denetim Standartları Genelgesi. URL: http://net.sgk.intra/wps/wcm/myconnect/8ec51c23-7eb4-40d9-bfe5-54c786a1ffe7/2013_05.pdf?MOD=AJPERES, Son Erişim Tarihi: 09.05.2019
- Kabak M., Sağlam F., Aktaş A., (2017). Farklı uzaklık hesaplama yaklaşımlarının topsıs üzerinde kullanılabilirliğinin incelenmesi, *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 32(1), 35-43.
- Karypis G., Han E., ve Kumar V. (1999). CHAMELEON: A hierarchical clustering algorithm using dynamic modeling. *Computer*, 32(8), 68-75.
- Kayısoğlu, B. (2014). *Maliyete Duyarlı Veri Madenciliği Yöntemleri İle Akciğer Kanseri Sınıflandırılması*, Yüksek Lisans Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Kayseri.

- Kaufman L., Rousseeuw P. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons, 68.
- Khan M., Ding, Q., Perrizo W. (2002). *K-Nearest neighbor classification on spatial data streams using p-trees*, Preceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, 517-528.
- Kıran, Z., B. (2010). *Lojistik Regresyon Ve Cart Analizi Teknikleriyle Sosyal Güvenlik Kurumu İlaç Provizyon Sistemi Verileri Üzerinde Bir Uygulama*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Kocabaş, F., M. (2010). *Veri Madenciliği Süreci ve Gerçek Bir Veri Üzerine Uygulanması*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Larose, D., T. (2005). *Discovering Knowledge In Data: an introduction to data mining*. New Jersey: John Wiley&Sons, 109,110.
- MacQueen J. (1967). *Some methods for classification and analysis of multivariate observations*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 281-297.
- Ngai, E. W. T., Xiu, Li, Chau, D. C. K. (2009). Application of data minning techniques in customer relationship management: A literatüre review and classification. *Expert Systems with Applications*, 36, 2592-2602.
- Orhan, U., Hekim M., Ozer M. (2011). EEG Signals Classification Using The K-Means Clustering And A Multilayer Perceptron Neural Network Model, *Expert Systems with Applications*, 38, 3475–13481.
- Özkan, Y. (2008). *Veri Madenciliği Yöntemleri (Birinci Basım)*. İstanbul: Papatya Yayıncılık Eğitim, 51, 55, 75, 77, 158.
- Pehlivan, G. (2006). *CHAID Analizi ve Bir Uygulama*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Raymond T., Han J. (1994). *Efficient and effective clustering methods for spatial data mining*. Proceedings of the 20th International Conference on Very Large Data Bases, 144-155.
- Shafer, J., Agrawal, R., Mehta, M. (1996). *SPRINT: A scalable parallel classiffier for data mining*. Proceedings of the 22th International Conference on Very Large Data Bases, 544-555.
- Sibson R. (1973). SLINK: An optimally efficient algorithm for the single link cluster method. *The Computer Journal*, 16(1), 30-34.
- Silahtaroglu, G. (2008). *Kavram ve Algoritmalarıyla Temel Veri Madenciliği (Birinci Basım)*. İstanbul: Papatya Yayıncılık Eğitim, 45, 59, 61, 70, 83, 99.
- Tan B., Steinbach M., Kumar V. (2006). *Introduction to Data Mining*. Boston: Pearson Addison-Wesley, 71, 72.

Zhang, S., Zhang C., Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, 17(5-6), 375-381.



ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : YAŞASINOĞLU, Nihal
 Uyuğu : T.C.
 Doğum tarihi ve yeri : 17.02.1987, Ankara
 Medeni hali : Evli
 Telefon : 0 (506) 631 49 48
 Faks : 0 (312) 435 94 92
 e-mail : mert_nihal@hotmail.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / İstatistik Bölümü	Devam ediyor
Lisans	Gazi Üniversitesi / İstatistik Bölümü	2010
Lise	Dr. Şerafettin Tombuloğlu Süper Lisesi	2004

İş Deneyimi

Yıl	Yer	Görev
2013-Halen	T.Vakıflar Bankası T.A.O	Uzman

Yabancı Dil

İngilizce

Yayınlar

Ebegil M., Yaşasinoğlu N., Yaşasinoğlu H. (2019). *Türkiye’de Muvazaalı Boşananların Veri Madenciliği Tekniklerinden Cart Algoritması İle Sınıflandırılması*, 1.Uluslararası 23 Nisan Multidisipliner Çalışmalar Kongresi, Ankara, Mayıs 2019.

Hobiler

Yüzme, Doğa, Rafting, Bisiklet, Film-Dizi



GAZİ GELECEKTİR..