



**T.C. İSTANBUL TİCARET
ÜNİVERSİTESİ**

**YAŞAM MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN SINIFLAMA VE
REGRESYON AĞACI İLE BELİRLENMESİ**

YASEMİN BAHAR YÜCEL

Danışman

YRD. DOÇ. DR SEDA BAĞDATLI KALKAN

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI
İSTANBUL - 2017**

KABUL VE ONAY SAYFASI

Yasemin Bahar YÜCEL tarafından hazırlanan "Yaşam Memnuniyetini Etkileyen Faktörlerin Sınıflama ve Regresyon Ağacı İle Belirlenmesi" adlı tez çalışması ~~20/07/2017~~ tarihinde aşağıdaki jüri üyeleri önünde başarı ile savunularak, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman

Yrd. Doç. Dr Seda Bağdatlı Kalkan
İstanbul Ticaret Üniversitesi



Jüri Üyesi

Prof. Dr. Ünal Halit ÖZDEN
İstanbul Ticaret Üniversitesi



Jüri Üyesi

Prof. Dr. Dilek ALTAŞ
İstanbul Ticaret Üniversitesi



Onay Tarihi: 15.08.2017



Yrd. Doç. Dr. Berk Ayvaz
Enstitü Md. V.

AKADEMİK VE ETİK KURALLARA UYGUNLUK BEYANI

İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

26.05.2017

YASEMİN BAHAR YÜCEL



İÇİNDEKİLER

Sayfa

İÇİNDEKİLER	ii
ÖZET	iii
ABSTRACT	iv
TEŞEKKÜR	v
ŞEKİLLER.....	vi
TABLolar	vii
SİMGE VE KISALTMALAR	viii
GİRİŞ	1
1 KARAR AĞAÇLARI	3
1.1 Karar Ağaçlarında Kullanılan Algoritmalar	6
1.1.1 CART Algoritması	7
1.1.2 CHAID Algoritması	8
2 SINIFLAMA VE REGRESYON AĞAÇLARI	10
2.1 Sınıflama Ağaçları.....	12
2.1.1 Gini Ayırma Kriteri	15
2.1.2 Twoing Ayırma Kriteri	16
2.1.3 Gini ve Twoing Ayırma Kriterlerinin Karşılaştırılması	17
2.1.4 Sınıflama Ağaçlarında Maksimum Boyuttaki Ağacın Oluşturulması ve Ağacının Büyüme Sürecinin Durdurulması	17
2.2 Regresyon Ağaçları	19
2.2.1 En Küçük Kareli Sapma (LSD)	20
2.2.2 Regresyon Ağaçlarında Maksimum Boyuttaki Ağacın Oluşturulması ve Büyüme Sürecinin Durdurulması	21
2.3 Sınıflama ve Regresyon Ağaçlarında Optimum Ağacın Belirlenmesi.....	22
3 YAŞAM MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN BELİRLENMESİ	24
3.1 Yaşam Memnuniyeti ve Mutluluk Kavramı	24
3.2 Uygulama.....	28
3.2.1 Değişkenlere ait Tanımlayıcı İstatistikler	29
3.2.2 CART Algoritması İle Oluşturulan Sınıflama Ağaçlarına Ait Bulgular	36
3.2.3 CHAID Algoritması İle Oluşturulan Sınıflama Ağaçlarına Ait Bulgular	44
3.2.4 CART Algoritması İle Oluşturulan Regresyon Ağaçlarına Ait Bulgular	51
3.2.5 CHAID Algoritması İle Oluşturulan Regresyon Ağaçlarına Ait Bulgular	57
4 SONUÇ.....	65
Kaynakça.....	68

ÖZET

Yüksek Lisans Tezi

YAŞAM MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN SINIFLAMA VE REGRESYON AĞACI İLE BELİRLENMESİ

Yasemin Bahar Yücel

İstanbul Ticaret Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Yrd.Doç.Dr. Seda Bağdatlı Kalkan
2017, 72 sayfa

Sınıflama ve Regresyon Ağaçları (SRA) varsayım gerektirmeyen istatistiksel tekniklerdir. SRA, bağımlı değişkenin kategorik olması durumunda sınıflama ağacı, bağımlı değişkenin sürekli olması durumunda ise regresyon ağacı ismini almaktadır. SRA; bağımsız değişkenler ve bağımlı değişken arasındaki anlamlı ilişkilerin modelini ağaç şeklindeki göstermektedir. Kolay yorumlanması, büyük veri setlerine uygulanması ve varsayım gerektirmemesinden dolayı son zamanlarda sıkça kullanılan tekniklerdir. Sınıflama veya Regresyon Ağacı oluşturan; Classification and Regression Tree (CART) ve Chi-square Automated Interaction Detection (CHAID) algoritmalarının en önemli özellikleri, sürekli ve kategorik verileri aynı anda modele dahil edebilmesi, bağımlı değişkenler üzerinde etkili olan bağımsız değişken(ler)i bir ağaç diyagramı üzerinde kolayca anlaşılabilir gösterebilmesi olarak özetlenebilmektedir.

Bu çalışmanın amacı, Türkiye'deki kişilerin mutluluk düzeylerini etkileyen faktörlerin sınıflama ve regresyon ağaçları ile belirlenmesidir. Hem sınıflama hem de regresyon ağaçlarında CART ve CHAID algoritmaları kullanılarak bu faktörlerin belirlenmesi amaçlanmıştır. Ayrıca, sınıflama ve regresyon ağaçları için farklı başlangıç ve test verileri kullanılarak ağaçlar oluşturulmuş, aralarındaki farklılıklar incelenmiş ve sonuçlar karşılaştırmalı olarak yorumlanmıştır.

Anahtar Kelimeler: Chi-square Automated Interaction Detection, Classification and Regression Tree, sınıflama ve regresyon ağaçları, yaşam memnuniyeti

ABSTRACT

M.Sc. Thesis

THE DETERMINATION OF FACTORS THAT AFFECT LIFE SATISFACTION WITH CLASSIFICATION AND REGRESSION TREES

Yasemin Bahar Yücel

Istanbul Commerce University
Graduate School of Applied and Natural Sciences
Department of Statistic

Supervisor: Yrd. Doc. Dr. Seda Bağdatlı Kalkan
2017, 72 pages

Classification and regression trees (CART) are techniques which don't require hypothesis. CART is named "classification tree" when the dependent variable is categorical, and it is named "regression tree" when the dependent variable is continuous. CART shows the model of the significant relations between dependent and independent variables as trees. Since it is easy to interpret, can be implemented to large data sets, and doesn't require hypothesis; it is often used recently. The most important features of Classification and Regression Tree (CART) and Chi-square Automated Interaction Detection (CHAID) algorithms are to be able to include continuous and categorical data at the same time, and to be able to show the independent variables that are effective on the dependent variables on a tree diagram which is very easy to understand.

The aim of this study is to determine the factors that affect the happiness levels of people living in Turkey with classification and regression trees. It is aimed to determine these factors, by using CART and CHAID algorithms on both classification and regression trees. Besides, trees are composed by using different beginning and test data for classification and regression trees; the differences between them are examined and the results are interpreted comparatively.

Keywords : Chi-square Automated Interaction Detection, Classification and Regression Tree, Life satisfaction

TEŞEKKÜR

Çalışmalarım boyunca bana faydalı olabilmek için büyük bir ilgiyle elinden gelenin fazlasını sunan, bilgi ve tecrübelerini benimle paylaşan, katkılarıyla beni yönlendiren, karşılaştığım her sorunda yanına çekinmeden gidebildiğim ve yardımlarıyla sorunları aşmamda yardımcı olan, ilgisini, sabrını ve güler yüzünü benden esirgemeyen, tez danışmanım ve çok kıymetli hocam Sayın Yrd. Doç. Dr. Seda BAĞDATLI KALKAN'a en içten minnet ve teşekkürlerimi sunarım.

Hem lisans hem de yüksek lisans öğrenimimde kendisine hayran olduğum ve öğrencisi olmaktan büyük onur duyduğum, değerli görüşleri ile gelecekteki hayatıma ışık tutan, alçak gönüllülüğü ve saygıdeğer kişiliğiyle herkesin sevgi ve saygısını kazanmış çok değerli hocam Sayın Prof. Dr. Münevver TURANLI'ya en içten teşekkürü borç bilirim.

Öğrenim hayatım boyunca değerli zamanlarını, bilgi ve tecrübelerini hiçbir zaman benden esirgememiş olan, eğitimciden ziyade aile yakınlığı kurabildiğim, bana araştırmayı ve sorgulamayı öğreten ve geleceğime ışık tutan saygıdeğer hocalarım Sayın Prof. Dr. Ünal Halit ÖZDEN ve Sayın Doç. Dr. Özlem Deniz BAŞAR'a en içten teşekkürlerimi sunarım.

Tez savunması esnasında yapıcı eleştirileriyle katkı sağlayan saygıdeğer hocam Sayın Prof. Dr. Dilek ALTAŞ'a teşekkürü bir borç bilirim.

Bütün öğrenim hayatım boyunca yaşamış olduğum sıkıntılarda ve yoğun dönemimde benden sabırlarını, ilgilerini ve sevgilerini esirgemeyerek güç veren ve bana inanan sevgili anne ve babama, aileme daha sonra yardımlarıyla her zaman yanımda olan sevgili arkadaşlarım Ezgi Senem KUL ve Fatma Yasemin AVCİL'e teşekkürlerimi sunarım.

Yasemin Bahar YÜCEL
İSTANBUL, 2017

ŞEKİLLER

Sayfa

Şekil 1 Sınıflama ve Regresyon Ağacı Düğümleri.....	11
Şekil 2 İkili Sınıflama Ağacı Alt Küme Örneği	13
Şekil 3 Katılımcıların Yaş Dağılımı	29
Şekil 4 Katılımcıların Cinsiyet Dağılımı.....	29
Şekil 5 Katılımcıların Medeni Durum Dağılımı	30
Şekil 6 Katılımcıların Çalışma Durum Dağılımı.....	30
Şekil 7 Katılımcıların Mutluluk Düzeyi Dağılımı	31
Şekil 8 Katılımcıların En Çok Kimin Mutlu Ettiğinin Dağılımı	31
Şekil 9 Katılımcıların En Çok Neyin Mutlu Ettiğinin Dağılımı	32
Şekil 10 Katılımcıların Geleceklerinden Umutlu Olma Dağılımı	33
Şekil 11 Katılımcıların Ülkenin Refah Düzeyine Göre Verdikleri Puan Dağılımı	34
Şekil 12 Katılımcıların "Ülkenin en önemli sorunu nedir?" Grüşü.....	36
Şekil 13 CART Algoritması ile Oluşturulan 1. Sınıflama Ağacı.....	37
Şekil 14 CART Algoritması İle Oluşturulan 2. Sınıflama Ağacı	39
Şekil 15 CART Algoritması İle Oluşturulan 3. Sınıflama Ağacı	42
Şekil 16 CHAID Algoritması İle Oluşturulan 1. Sınıflama Ağacı.....	45
Şekil 17 CHAID Algoritması İle Oluşturulan 2. Sınıflama Ağacı.....	47
Şekil 18 CHAID Algoritması İle Oluşturulan 3. Sınıflama Ağacı.....	49
Şekil 19 CART Algoritması İle Oluşturulan 1. Regresyon Ağacı.....	51
Şekil 20 CART Algoritması İle Oluşturulan 2. Regresyon Ağacı	53
Şekil 21 CART Algoritması İle Oluşturulan 3. Regresyon Ağacı	56
Şekil 22 CHAID Algoritması İle Oluşturulan 1. Regresyon Ağacı	58
Şekil 23 CHAID Algoritması İle Oluşturulan 2. Regresyon Ağacı	60
Şekil 24 CHAID Algoritması İle Oluşturulan 3. Regresyon Ağacı	62

TABLULAR

	Sayfa
Tablo 1.Karar Ağaçları Model Teması.....	4
Tablo 2.Karar Ağaçlarında Kullanılan Sınıflama ve Regresyon Ağaçları Algoritmaları .	6
Tablo 3. 2003-2016 Yılları Arası Mutluluk Düzeyi Değerleri.....	26
Tablo 4. Katılımcıların Mutlulukla İlişkilendirilen Konulardaki Memnuniyet Düzeyleri	32
Tablo 5. Katılımcıların Bugüne Kadar Aldıkları Eğitimden Memnuniyet Düzeyleri	33
Tablo 6.Katılımcıların Çeşitli Konulara Olan İlgisi Düzeylerinin Dağılımı	34
Tablo 7. Katılımcıların "Ülkenin en önemli sorunu nedir?" Görüşü	35



SİMGE VE KISALTMALAR

AID	Automatic Interaction Detector
CART	Classification and Regression Trees
CHAID	Chi-Squared Automatic Interaction Detector
LSD	Least Squared Deviation
SRA	Sınıflama ve Regresyon Ağaçları
TÜİK	Türkiye İstatistik Kurumu



GİRİŞ

Sınıflama ve Regresyon Ağaçları(SRA) verilerin görsel olarak sunulması, aralarındaki etkileşimin belirlenebilmesi ve bu etkileşime göre tahminlerin yapılabilmesi için sıkça kullanılan karar ağacı yöntemlerinden biridir. Verilerin analizi için istatistiksel varsayımları yerine getirebilmek oldukça güçtür. İstatistiksel varsayımların sağlanamadığı durumlar için bir çok teknik geliştirilmiştir. Bu tekniklerinden biri olan SRA; hiçbir varsayım gerektirmeden sonuçların ağaç modeli olarak görselleştirilmesinden dolayı herkes tarafından kolaylıkla anlaşılabilir ve yorumlanabilir. SRA; bağımlı değişkenin kategorik olması durumunda sınıflama ağacı, sürekli olması durumunda regresyon ağacı olarak adlandırılmaktadır.

SRA'da veri dallanan bir ağaç şeklinde hiyerarşik bir düzende sunulur. SRA'nın temel amacı bağımsız değişkenler matrisini homojen alt gruplara ayırmaktır. Alt grupların oluşturulmasında veri dallanan bir ağaç şeklinde hiyerarşik bir düzende sunulur. Ağaç içindeki ara düğümlerde en iyi ayrımı yapmış olan bağımlı değişkenler gösterilir. Bu düğümlerin dallarında ayırıcı bağımlı değişkenlerin kritik değeri verilir. Yapraklar bağımlı değişkenin değerlerini gösterir. Kök düğüm (ilk düğüm noktası) noktasından itibaren yapraklara kadar (en son düğüm) hatlar bulunmaktadır. Bu hatlar boyunca sınıflar arası ayrımın maksimize edildiği ve her sınıfın içindeki varyasyonun minimize edildiği ayrımların kuralları gösterilmektedir (Özkan K. , 2012).

CART ve CHAID algoritmaları SRA yönteminde kullanılan alternatif algoritmalarındandır. Veri seti çok karmaşık olsa bile bağımlı değişkeni etkileyen bağımsız değişkenleri ve bu değişkenlerin önemini ağaç şeklinde ortaya koyarlar.

Bu çalışmanın amacı; Sınıflama ve Regresyon Ağaçlarında CART ve CHAID algoritmaları ile 2015 yılı Türkiye geneli mutluluk düzeylerini etkileyen faktörlerin belirlenmesidir. Hem sınıflama ağacı için hem de regresyon ağacı için farklı başlangıç ve test verileri kullanılarak ağaçlar oluşturulmuş, aralarındaki farklılıklar incelenmiş ve sonuçlar karşılaştırılmalı olarak yorumlanmıştır.

Çalışmanın birinci bölümünde, SRA'nın karar ağaçlarının bir yöntemi olmasından dolayı karar ağaçlarından kısaca bahsedilmiş ve karar ağaçlarının uygulama alanları belirtilmiştir. Daha sonra karar ağaçları analizinde SRA yönteminde kullanılan algoritmalar ile ilgili kısaca bilgi verilip bunlardan çalışmada kullanılan CART ve CHAID algoritmaları anlatılmıştır.

Çalışmanın ikinci bölümünde SRA genel olarak anlatılmış daha sonra sınıflama ağaçları ve regresyon ağaçları teorik olarak açıklanmış ve optimum ağacın belirlenmesi süreci anlatılmıştır.

Çalışmanın üçüncü bölümünde; Türkiyedeki kişilerin mutluluk düzeylerini etkileyen faktörler hem sınıflama hem de regresyon ağacı için CART ve CHAID algoritmaları kullanılarak belirlenmeye çalışılmıştır. Uygulamada Türkiye İstatistik Kurumu (TÜİK)'nin 2015 yılında Türkiye genelinde 9397 kişiye uyguladığı "Yaşam Memnuniyeti Anketi" verileri kullanılmıştır. İlk olarak yaşam memnuniyeti, mutluluk kavramı ve Türkiye'de bu konu ile ilgili yapılmış çalışmalardan bahsedilmiştir. Daha sonra uygulamaya geçilmiştir. Uygulamanın birinci aşamasında, mutluluk düzeylerini etkileyen faktörlerin CART ve CHAID algoritmaları kullanılarak sınıflama ağaçları ile belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda sırasıyla, model verisinin %30, %50 ve %70'lik kısmı analize dahil edilerek sınıflama ağaçları oluşturulmuş ve oluşturulan ağaçlar karşılaştırmalı olarak yorumlanmıştır. Uygulamanın ikinci aşamasında memnuniyet puanı değişkeni oluşturulmuş ve bu değişkeni etkileyen faktörlerin CART ve CHAID algoritmaları kullanılarak regresyon ağaçları ile belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda sırayla, model verisinin %30, %50 ve %70'lik kısmı analize dahil edilerek regresyon ağaçları oluşturulmuş ve oluşturulan ağaçlar karşılaştırmalı olarak yorumlanmıştır.

1 KARAR AĞAÇLARI

Karar ağaçları bir tüme varım yöntemidir ve bilgi teorisi ilkelerine dayanır. Karar ağaçları parametrik olmayan istatistiksel yöntemlerde kullanılan bir analiz yöntemidir. Parametrik yöntemler, incelenen örneklemin sağladığı varsayımlara dayanır. Bu varsayımlar kitlenin normal dağılmış olması, örneklenen kitlenin varyanslarının eşit olması, gözlemlerin bağımsız olması yani otokorelasyon sorununun olmaması ve aralıklı bir ölçekle ölçülmüş olmasıdır. Bu varsayımların her zaman tam anlamıyla sağlanamaması zordur. Varsayımların sağlanamadığı durumlarda parametrik olmayan istatistiksel yöntemlere başvurulur.

Karar ağaçları; varsayımları sağlama zorunluluğu olmaması, düşük maliyetli olması, anlaşılmasının, yorumlanmasının ve veri tabanları ile entegrasyonun kolaylığı, güvenilirliklerinin iyi olması gibi nedenlerden ötürü en yaygın kullanılan sınıflandırma tekniklerinden biridir.

Bu yöntemin tercih edilmesinin en önemli sebebi, aynı anda analize giren bağımsız değişkenlerin hangi seviyede bağımlı değişken ile ilgili olacağı ve bağımsız değişkenlerin seviyelerine göre nasıl bir sınıflama olduğunu ortaya koymasındadır. Hedef, gerçekleşen sınıflama sonucunda bir karara varmaktır. Bu tekniğin en büyük avantajlarından birisi de görsel olarak bir ağaç yapısı sunmasıdır (Baran & Ata, 2014).

Karar ağaçlarının avantajları yanında dezavantajları da bulunmaktadır. Genellikle karar vermek karar ağacı yaklaşımı için ciddi bir sorun oluşturur. Sorun şundan kaynaklanır; ağaç, ayrılan bölümleriyle genişledikçe, yapılan sınıflandırma sonucunda bölüm düğümlerinde daha az bilgi kalacaktır. Karar ağacı veriyi çok fazla sayıda takıma böler. Bu takımlar daha spesifik bir hal aldıkça küçülmeye başlar. İncelenmesi gereken farklı durumların sayısı arttıkça, eğitim kümelerinin her biri daha da küçülür (Altaş & Gülpınar, 2012). Rakamlardaki azalma yüzünden sınıflandırmanın doğru bir şekilde gösterilmesinde daha az güvenilirlik söz konusu olur. Karar ağacı çok fazla sayıda küçük dallardan oluşturulursa bu düğümlerin içerisinde herhangi doğru olabilecek bir istatistiksel incelemeden başarıyla çıkabilecek kuralların bulunması iyimser bir olasılıktır. Çünkü, genellikle bu dallardan çıkacak olan her bir düğüm, olası bütün sınıfların küçük oranlarını içerecektir. Bu da uygulamada sorunlara yol açabilmektedir (Seidman, 2001).

Karar ağacı algoritması kullanıcıların, karışık ve bilinmeyen verilerin üstesinden gelmesini ve kolayca yorumlanmasını sağlayan modeller oluşturur. Ağacın üst düğümü kök düğüm olarak adlandırılır; içindeki düğümler değişkenlerin değerleri üzerine belirlenmiş olan testleri temsil ederler; dallar testler üzerindeki farklı sonuçları ve yaprak düğümleri örneklerin sınıfını temsil eder. Karar ağacı süreci kök düğümden başlayarak, yukarıdan aşağıya doğru ardışık düğümler takip edilerek yaprağa ulaşıncaya kadar devam eden bir süreçtir (Soner, 2015).

Tablo 1.Karar Ağaçları Model Teması

Düğüm Adı	Tanımı	Ağaçta Bulunduğu Yer
Kök Düğüm	Ana düğümdür. Verideki ilgili bağımsız değişkene ait bütün gözlemleri kapsar.	Başta
Ara Düğüm	Karar düğümüdür. Yaprak değildir.İlgili değişkenin test sonucuna göre yeniden bölünür.Bu yüzden uçta bulunmayan düğümdür.	Ortada
Uç Düğüm	Yaprak düğümdür. Bölünme yapmaz.Sınıfa atama olur.	Sonda

Kaynak : (Saitoğlu, 2015)

Karar ağaçları genellikle iki aşamada oluşturulurlar. Büyüme olduğu zaman bu algoritma her bir düğümde sınıflar arasındaki en iyi özellik ayırt ediciyi (veri alt seti) ortaya çıkarır ve daha sonra o özelliğe dayalı olarak bu verileri iki yeni düğüm halinde bölümlere ayırır. Bu, her bir tabaka için bir sınıf tahsis edilinceye kadar ortaya çıkan veri alt setine tekrar tekrar uygulanır. İndirgemenin ikinci aşaması en iyi dengeye erişebilmek için ağacın en az yararlı dallarını kesmek suretiyle işletilir. Daha basit bir model genellikle daha sağlamdır. Yani yeni veriler hakkında daha doğru sonuçlar ortaya çıkarır. Nihai ağaç etiketlenmiş birkaç bölgedeki bu özel alanı bölümlere ayırır (Nisbet, Elder , & Miner, 2009).

Başlangıç veri kümesi içinde geçmiş bilgileri barındırır ve bütün gözlemlerin önceden hangi sınıflara atanmış olduğunu bilir. Böylelikle geçmiş veri kümesine göre karar ağacı oluşturulur ve yeni veri kümesi de bu ağaçtaki bilgilere göre sınıflandırılır (Timofeev, 2004).

Karar ağaçlarıyla ilgili genel tanım ve kavramlar bu şekildedir. Fakat karar ağaçları ile sınıflamanın daha iyi anlaşılması için matematiksel gösterimler de aşağıda kısaca açıklanmıştır.

Buna göre \mathbf{x}^1 ; ölçüm vektörü (measurement vektor) olarak adlandırılır ve bir gözlem için ölçülmek istenen değişkenlerinin ölçüm sonuçlarını içinde barındırır. Yani p adet değişkenden oluşan bir gözlemin vektörü olarak düşünülebilir. Buna göre ölçüm vektörü şu şekilde gösterilir:

$$\mathbf{x}^1 = (x_1, x_2, \dots, x_p) \quad [1.1]$$

\mathbf{X} ; ise ölçüm uzayı (measurement space) olarak adlandırılır ve bütün gözlemlerin ölçüm sonuçlarını yani ölçüm vektörlerini içinde barındıran bir matristir. Ölçüm uzayı içinde hangi ölçümler önemliyse onun ölçüm vektörünü taşır. Bu yüzden birçok ölçüm uzayı matrisi oluşturulabilir. Buna göre (nxp) boyutlu \mathbf{X} matrisi [1.2] deki gibi gösterilir:

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} \quad [1.2]$$

Burada $g=1, \dots, n$ ve $l=1, \dots, p$ olmak üzere; n gözlem sayısını, p ise bağımsız değişken sayısını göstermektedir. Buna göre $\mathbf{X}_{n \times p}$ matrisinde ilk indis gözlem numarasını, ikinci indis ise değişken numarasını gösterir.

Oluşturulan karar ağacı \mathbf{T} ile ifade edilmek üzere, $\tilde{\mathbf{T}}$ ağaçtaki uç düğümlerin oluşturmuş olduğu bir kümeyi ifade etmek için kullanılacaktır. Gözlemlerin J adet sınıfa atandığını varsaydığımızda sınıf sayısı $1, 2, \dots, J$ adet olmak üzere, C sınıf kümesi olarak adlandırılır ve;

$$C = \{1, 2, \dots, J\} \quad [1.3]$$

ile gösterilir. Burada C 'ler aslında bağımlı değişken değeridir. Bağımlı değişkende kaç adet sınıf yani kategori olduğunu ifade eder (Saitoğlu, 2015).

Bir sınıf üyeliğini tespit ederken izlenecek yol $\mathbf{x} \in \mathbf{X}$ olmak üzere her bir \mathbf{x} vektörünü (yani her bir gözlemi), 1 'den J 'ye kadar olan bu sınıflardan birine atamaktır. Buna göre bir sınıflayıcı ya da sınıflama kuralı (a classifier or classification rule), $d(\mathbf{x})$ ile gösterilen bir fonksiyon ile tanımlanır. Her problemde sınıflayıcının yani sınıflama kuralının görevi bir gözlemin ya da nesnenin hangi sınıfta yer alacağını sistematik bir şekilde tahmin etmektir.

Buna göre $d(\mathbf{x})$ fonksiyonu, \mathbf{X} matrisi ve dolayısıyla \mathbf{x} vektörü üzerinde tanımlı bir fonksiyondur ve 1 'den J 'ye kadar olan sınıflardan birine eşittir $\{d(\mathbf{x}) = j\}$. Yani sınıflandırıcının $[d(\mathbf{x})]$ asıl amacı, \mathbf{X} matrisinin bir alt kümesini olarak tanımlanan A_j 'yi oluşturmaktır.

$$A_j = \{\mathbf{x}; d(\mathbf{x}) = j\} \quad [1.4]$$

[1.4] nolu eşitliğe göre \mathbf{x} vektörü, $d(\mathbf{x})$ olarak tanımlanan bir sınıflandırıcı fonksiyonu aracılığıyla j sınıfına atanmakta ve A_j denilen yani j sınıfını içinde barındıran bir alt küme oluşturmaktadır. Böylelikle her bir gözlemin sınıflara ataması gerçekleştirilir. Bu durum bize başlangıç veri kümesinin nasıl bir yol izleyerek alt parçaya ayrıldığını gösterir. Çünkü A_1, \dots, A_j 'ler \mathbf{X} matrisinin alt kümeleridir ve $\mathbf{X} = \bigcup_j A_j$ olup, her \mathbf{x} vektörü tahmini bir j sınıfına atanmaya çalışılır. Atandığı sınıf içinde $\mathbf{x} \in A_j$ 'dir (Breiman L. , Freidman, Olshen, & Stone, 1993).

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı alanlar ve belli başlı uygulamalar; belirli bir sınıfın üyesi olacak elemanların belirlenmesi (Segmentation), çeşitli vakaların yüksek, orta, düşük risk grupları gibi kategorilere ayrılması (Stratification), sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması, kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikli değişkenlere dönüştürülmesi, parametrik modellerin kurulmasında kullanılmak üzere çok sayıdaki değişkenden en önemlilerinin seçilmesi, gelecekteki olayların tahmin edilmesi için kurallar oluşturulması, bireylerin kredi geçmişlerini kullanarak kredi vakalarının verilmesi, tıbbi gözlem verilerinden yararlanarak en etkin kararların verilmesi, geçmişte işletmeye en faydalı olan bireylerin özelliklerini kullanarak işe alma

süreçlerinin belirlenmesi, satışları etkileyen değişkenlerin belirlenmesi, üretim verilerinin incelenmesiyle ürün hatalarına yol açan değişkenlerin belirlenmesi olarak özetlenebilir (Kıran, 2010).

1.1 Karar Ağaçlarında Kullanılan Algoritmalar

Karar ağaçlarına dayalı olarak geliştirilen bir çok algoritma vardır. Bu algoritmalar kök, düğüm ve dallanma kriterleri seçimlerinde izledikleri yol açısından birbirlerinden ayrılırlar. Karar Ağaçlarında kullanılan bu algoritmalar bağımlı ve bağımsız değişken türüne göre Sınıflama yada Regresyon Ağacı olarak ayrılmaktadır. Karar ağaçları, bağımlı değişkeni tahmin edici değişkenlere göre ayırma mantığına dayansa da, bünyesinde değişik amaçlara hizmet eden birbirinden farklı algoritmalara sahiptir. Analizde kullanılan verinin türü ve amacına göre uygulanacak algoritmaya karar verilir. Tablo 2’de Karar Ağaçlarında kullanılan algoritmaların bir kısmı özetlenmeye çalışılmıştır.

Tablo 2. Karar Ağaçlarında Kullanılan Sınıflama ve Regresyon Ağaçları Algoritmaları

Algoritma Adı	Veri Tipi	Özellikleri	Ağaç Dallanma Kriteri
CART Algoritması	Bağımlı değişken kategorik (ordinal / nominal) veya sürekli olabilir. Bağımsız değişken için ayrımı yoktur.	Son veya uç olmayan her bir düğümde iki adet dal bulunmaktadır. Budama işlemi ağacın karmaşıklık ölçüsüne dayanır. Sınıflama ve regresyonu destekleyici bir yapıdadır. Sürekli hedef değişkenleri ile çalışır. Bağımlı değişken kategorik ise, sınıflama ağacı- sürekli ise regreyon ağacı oluşur.	Bağımlı değişken kategorik → gini / twoing kriteri, Bağımlı değişken sürekli → EKK yöntemi kullanılır.
CHAID Algoritması	Bağımlı değişken kategorik (ordinal / nominal) veya sürekli olmalı. Bağımsız değişken kategorik (nominal / ordinal) olmalıdır.	Dalların sayısı iki ile tahmin edicinin kategori sayısı arasında değişir. Ağaç çoklu alt gruplara ayrılır. Yani çoklu ağaçlar oluşturur. Bağımlı değişken kategorik ise, sınıflama ağacı- sürekli ise regreyon ağacı oluşur.	Bağımlı değişken sürekli → F testi- Bağımlı değişken nominal → pearson ki-kare testi, bağımlı değişken ordinal → en çok olabilirlik testi kullanılır.
C4.5- C5.0 Algoritması	Bağımlı değişken kategorik (ordinal / nominal), bağımsız değişkenler ise kategorik veya sürekli olabilir.	Her düğümden çıkan çoklu dallar ile ağaç oluşturur. Dalların sayısı tahmin edicinin kategori sayısına eşittir. Tek bir sınıflayıcı da birden çok karar ağacını birleştirir. Bağımlı değişken kategorik olduğu için sınıflama ağacı oluşur.	Kazanç ölçütü kriterini kullanır. Budama işlemi her yapraktaki hata oranına dayanır.
ID3 Algoritması	Bağımlı ve bağımsız değişkenler kategorik (ordinal / nominal) olmalıdır.	Her düğümden çıkan çoklu dallar ile ağaç oluşturur. Dalların sayısı tahmin edicinin kategori sayısına eşittir. Tek bir sınıflayıcı da birden çok karar ağacını birleştirir. Bağımlı değişken kategorik olduğu için sınıflama ağacı oluşur.	Kazanç ölçütü kriterini kullanır. Budama işlemi her yapraktaki hata oranına dayanır

MARS Algoritması	Bağımlı değişken ikili (binary) veya sürekli olabilir. Bağımsız değişken için ayırım yoktur.	Ağaçta sadece ikili bölünmeler olur. Yani ikili ağaçlar oluşturur. Bağımlı değişken kategorik ise, sınıflama ağacı, sürekli ise regresyon ağacı oluşur.	İleriye ve geriye doğru adım algoritmalarını kullanır.
AID Algoritması	Bağımlı değişken nicel (aralıklı/oransal) olmalıdır. Bağımsız değişken kategorik (nominal /ordinal) olmalıdır.	Ağaçta sadece ikili bölünmeler olur. Yani ikili ağaçlar oluşturur. Bağımlı değişken nicel olduğu için regresyon ağacı oluşur.	Gruplar arası kareler toplamı kriterini kullanır.
SLIQ Algoritması	Bağımlı değişken kategorik (ordinal / nominal), bağımsız değişkenler kategorik veya sürekli olabilir.	Hızlı ölçeklenebilir bir sınıflayıcıdır. Hızlı ağaç budama algoritması mevcuttur.Ağaç çoklu alt gruplara ayrılır. Yani çoklu ağaçlar oluşturur. Bağımlı değişken kategorik olduğu için sınıflama ağacı oluşur.	Gini kriterini kullanır.
OUEST Algoritması	Bağımlı değişken kategorik (nominal),bağımsız değişkenler kategorik veya sürekli olabilir.	Bağımlı değişkeni sadece ikiye böler. Yani ikili ağaçlar oluşturur. Bağımlı değişken kategorik olduğu için sınıflama ağacı oluşur.	Kuadratik diskriminant analizini kullanır.
SPRINT Algoritması	Bağımlı değişken kategorik(ordinal / nominal), bağımsız değişkenler kategorik veya sürekli olabilir.	Büyük veri kümeleri için idealdir. Bölme işlemi tek bir niteliğin değerine dayanır. Tüm bellek sınırlamaları üzerinde nitelik listesi veri yapısı kullanarak işlem yapar.Ağaç çoklu alt gruplara ayrılır. Yani çoklu ağaçlar oluşturur. Bağımlı değişken kategorik olduğu için sınıflama ağacı oluşur.	Gini kriterini kullanır.

Kaynak : (Saitoğlu, 2015)

1.1.1 CART Algoritması

CART Algoritması, Morgan ve Sonquist'in AID (Automatic Interaction Detection) adlı karar ağacı algoritmasının devamı niteliğine Breiman ve diğerleri tarafından 1984 yılında önerilmiştir (Breiman L. , Freidman, Olshen, & Stone, 1984). Hem sayısal hem de nominal veri türlerini, girdi ve kestirimsel değişken olarak kabul edebilen CART algoritması, sınıflama ve regresyon problemlerinde bir çözüm olarak kullanılabilir (Sezer, Bozkır, Yağız, & Gökçeoğlu, 2010). Yöntemin asıl amacı; incelenmekte olan probleme yönelik tahmin yapısını ortaya çıkaran hatasız bir veri seti sınıflayıcısını oluşturmaktır. Sınıflamanın amacı ise, karakterize edilmiş ağaç sayesinde gelecekteki herhangi bir değer için hangi sınıfa düşeceğini belirlemektir. Bu amaçla, hangi değişkenlerin ya da değişkenler arası etkileşimin en iyi sonucu tahmin etmek için

gerekli olduğu belirlenir (Yohannes & Hoddinot, Classification and Regression Trees: An Introduction, 1999).

CART yöntemi ile ağaç oluşturma süreci, üç temel kural içermektedir. Bunlar sırasıyla; “örnekleme bölme kuralı (sample-splitting rule)”, “bölme uyum kriteri (goodness of split criteria)” ve “optimum ağacı seçme” kriteridir.

CART önceden belirlenmiş bölme kuralına ve düğüm bölme sürecinin her aşamasında uygulanan bölme uyum kriterine göre ağaç oluşturur. Ağaç oluşturma sürecinde CART başlangıç veri setini, bağımsız değişkene sorduğu sorulardan alacağı evet/hayır cevaplarına göre ikili alt parçaya (alt örnekleme) böler. Her soru sadece bir bağımsız değişken için sorulur. Alınan cevaba bağlı olarak veri seti sağ ve sol alt parçaya ayrılır (Yohannes & Webb, 1999). CART algoritmasının daha fazla dallanma potansiyeli olan bir algoritma olduğu söylenebilir. Bu gibi durumları önlemek için çalışılan veri setini öğrenme (Learning, Training) ve test (testing, validation) setlerine ayrılması, ebeveyn-yavru düğümde bulunması gereken birey sayısının ve ağaç derinliği (tree depth) sayısının dikkatli bir şekilde ayarlanması önerilebilir (Koç, 2016).

Ağaç yapısını kullanan algoritmalar incelendiğinde CART çok sayıda avantaja sahiptir. En büyük avantajı parametrik olmadığı için bağımsız değişken değerlerine ilişkin varsayım gerektirmemesidir. Bundan dolayı değişkenler sayısal değişkenler olabileceği gibi kategorik değişkenler de olabilir. Bu sayede araştırmacı analiz için dönüşüm işlemiyle uğraşmayacağı için zaman kazanmış olacaktır (Ersöz, Özseven, & Ersöz, 2017).

1.1.2 CHAID Algoritması

CHAID algoritması 1980 yılında Gordon V. Kass tarafından geliştirilen algoritma, bölmeleme veya ağaç oluşturma için oldukça etkili istatistiksel bir yöntemdir. CHAID istatistiksel bir testi dallanma kriteri olarak kullanarak, tüm bağımsız değişkenler için bir değerlendirme yapmaktadır. Bağımlı değişken değerlerine göre istatistiksel olarak homojenlik gösteren (farklılık göstermeyen) değerleri birleştirmekte, farklılık gösterenlere ise dokunmamaktadır. Daha sonra dallarında homojen değerlerin yer aldığı ağacın ilk dalını oluşturmak için en iyi bağımsız değişkeni seçmekte ve bu süreç ağaç tamamen büyüyene kadar yinelemeli bir şekilde devam etmektedir. Dallanma kriteri olarak kullanılacak istatistiksel test bağımlı değişkenin ölçme düzeyine göre farklılık göstermektedir. Eğer hedef değişken sürekli ise F testi, kategorik ise ki-kare (X^2) testi kullanılmaktadır. En küçük p değerine sahip değişken ile ağacın ilk dalı oluşturulmaktadır (Atılğan, 2011).

CHAID algoritması ile diğer algoritmalar arasındaki en önemli farklılık, diğer yöntemler ikili ağaçlar türetirken, CHAID algoritması çoklu ağaçlar türetmektedir (Albar Özbucak, 2014). Bu algoritma ile sürekli ve kategorik veriler aynı anda modele dahil edilebilmektedir (Kayri & Boysan, 2007). Başka bir ifadeyle bağımlı ve bağımsız değişkenlerin tümünün aynı tip ölçekle ölçülmüş olmasına gerek duyulmamaktadır (Koyuncu & Özgülbaş, 2008).

Bu sebeple CHAID algoritması parametrik ve parametrik olmayan ayrımını kaldırmakta ve yarı parametrik (semi- parametric) bir özellik taşımaktadır (Kayri & Boysan, 2007).

CHAID algoritması ile çok kategorili deęişkenlerin yer aldığı büyük bir veri kümesi, benzer kategoriler birleştirilerek ve önemli sayılan deęişkenlere göre bölünerek, indirgenir. Her bir bağımsız deęişken için kategorilerin anlamlı bir şekilde birleştirilmesinden sonra, bağımlı deęişkene göre kontenjans tabloları oluşturularak, Bonferroni p deęerleri ile χ^2 istatistikleri hesaplanır. Bağımsız deęişkenler birbirleri ile karşılaştırılıp en küçük Bonferroni p deęerine sahip deęişkenin kategorilerine göre, veriler tekrar alt gruba ayrılır (Silahtaroglu, 2013).

CHAID algoritmasında özellikle bağımsız deęişkenlerin, birbirleriyle olan ilişki ve etkileşimleri konu edildiği için χ^2 istatistiği kullanılmaktadır. Ki-kare test istatistiği, deęişkenler arasındaki bağımlılığı ele almaktadır (Kayri & Boysan, 2007). CHAID algoritması bağımlı deęişken ile bağımsız deęişkenler arasındaki ilişkiyi belirlemekle beraber, bağımsız deęişkenlerin birbirleriyle olan etkileşimlerini de ortaya koymaktadır (Kayri & Boysan, 2008). Burada dikkat edilmesi gereken husus ise kategoriler birleştirilirken kullanılan anlamlılık deęeridir (*p value*). Bunun için bağımlı deęişkenin türüne bakılır. Eđer bağımlı deęişken sürekli ise F testi kullanılır. Eđer bağımlı deęişken nominal kategorik bir deęer ise pearson ki-kare istatistiği kullanılır. Eđer ordinal kategorik ise en çok olabilirlik testi (likelihood ratio) kullanılır. Her birleştirilmiş kategori için de bulunan p deęerinin anlamlı olup olmadığına bakılır. Bu süreç anlamlı olmayan kategori birleşmesi bulununca son bulur (Rokach & Maimon).

2 SINIFLAMA VE REGRESYON AĞAÇLARI

SRA, varsayım gerektirmeyen, kategorik ya da sürekli bağımlı değişkenin sınıf üyeliğini tahmin eden ters ağaç şeklindeki modellerdir (Temel, 2004).

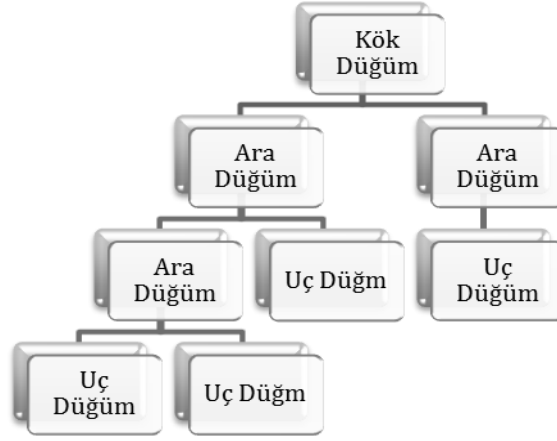
Kategorik ya da sürekli, bir ya da birden fazla bağımsız değişkenin kombinasyonları kullanılarak, homojen bölünmelerle, bağımlı değişkendeki değişimi ortaya çıkarmaya ve bağımlı değişkenin değerlerini tahmin etmeye yarayan ve görsel olarak ters ağaç şeklindeki modellere ağaç modelleri denir (Mertens, Nestler, & Huwe, 2002). Ağaç dalları bağımlı değişkeni hangi bağımsız değişken ya da değişkenlerin etkilediğini gösterir.

Ağaç modellerinin iki temel amacı vardır. Birinci amacı, basit yapıdaki verileri sistematik olarak modelde sunması yani sınıflamasıdır. İkinci amacı ise gözlenmeyen verilerin, oluşturulan modelde tahmin edilmesidir.

Regresyon analizlerinde ağaç kullanım yaklaşımı ilk olarak 1960'lı yıllarda Morgan ve Sonquist tarafından geliştirilen AID (Automatic Interaction Detection) yöntemine dayanmaktadır (Morgan & Sonquist, 1963). 1973 yıllarında ise Breiman ve Friedman birbirinden bağımsız olarak sınıflamada ağaç yöntemlerinin kullanılmasını yeniden keşfetmek üzere çalışmalar yapmıştır. Daha sonrasında Breiman ve Friedman çalışmalarını biraraya getirmiş ve Stone'un da katılımıyla yöntemin gelişmesine anlamlı katkılar sağlamıştır. Olshen ise medikal uygulamalarda ağaç yönteminin ilk kullanıcılarından olup, ekibin çalışmalarının teorik gelişmelerine katkılar sağlamıştır. Ağaçlarla ilgili bütün bu çalışmalar Friedman (Breiman L. , Freidman, Olshen, & Stone, 1984) tarafından birleştirilmiş ve 1980'li yıllarda CART (Classification and Regression Trees) adı verilen bu tekniğin doğmasını sağlamıştır (Sutton, 2005).

SRA yöntemi bağımlı değişkenin sürekli olması durumundaki kullanımı ile çoklu regresyon analizini, kategorik olması durumunda da lojistik regresyon analizini kapsamaktadır. Fakat SRA yöntemi parametrik regresyon analizinin sağlamak zorunda olduğu varsayımlardan sorumlu değildir. İncelenen veri seti üzerinde hiçbir varsayım gerektirmemesi nedeniyle, SRA bu tür istatistiksel sınıflama ve regresyon tekniklerine karşı güçlü bir alternatif olarak ortaya çıkmaktadır.

SRA'nın, aynı aile soyağacında olduğu gibi hiyerarşik bir yapısı vardır. Bu yüzden daha önce belirtilmiş olan kök, ara ve uç düğüm ifadeleri tıpkı aile soyağaçlarında kullanılan ata, ana, kardeş, çocuk ve torun gibi birçok kavramı içermektedir (Çölkesen, 2014).



Şekil 1 Sınıflama ve Regresyon Ağacı Düğümleri

Şekil 1’de görüldüğü üzere, 1. Düzeyde bulunan kök düğüm ata düğümdür. Bir düğüme doğrudan bağlı olan düğümler ise o düğümün çocuklarıdır. 2. Düzeyde bulunan 2 ara düğüm yani 2 tane çocuk düğüm vardır. 2. Düzeydeki ara düğümlerde 3. Düzeyde görülen toplam üç tane çocuk düğümleri bulunmaktadır. 3. Düzeyde bulunan düğümler 1. Düzeydeki kök düğümün torunlarıdır. Aynı zamanda aynı düğüme bağlı olan düğümler kardeş düğümlerdir. Yani 3. Düzeyde bulunan ara ve uç düğüm 2. Düzeydeki ara düğüme bağlı olduğu için kardeşlerdir. 4. Düzeydeki uç düğümler 3. Düzeydeki ara düğümün çocuklarıdır ve aynı ara düğüme bağlı olduğu için kardeşlerdir. Uç düğümler bölünemediği için 3. Düzeydeki uç düğümlerin çocukları bulunmamaktadır. Ayrıca 4. Düzeydeki uç düğümler hem 1. Düzeydeki kök düğüm olan ata düğümün hemde 2. Düzeydeki ara düğümün torunlarıdır.

SRA yöntemiyle bağımsız değişkenler ve bağımlı değişkenler arasındaki önemli ilişkilerin açılımını veren etkileşim modelinin diyagram (alt ağaçlar; node) biçimindeki gösterimi, bu tekniğin görsel bir üstünlüğü olarak da düşünülebilir. Sınıflama ve regresyon ağacı, sadece bağımlı değişken ile bağımsız değişken arasındaki ilişkinin yapısını araştırmakla kalmaz; aynı zamanda bağımsız değişkenlerin birbirleri ile olan etkileşimlerini de ortaya koymaya çalışmaktadır. SRA’nın sahip olduğu algoritma, benzerlik gösteren değişkenlerin aynı ağaç düğümünde toplanmasına dayalı olup, bütün oluşturduğu alt dalları bağımlı değişken olan kök düğüme bağlamayla son bulmaktadır (Kayri & Boysan, 2008).

SRA’yı oluşturabilmek için ağaca önce büyüme, daha sonra budama işlemi uygulanır (Küçüköğlü, 2010). SRA daha çok mühendislik, tıp, zooloji ve sanayi alanındaki analizlerde tercih edilmektedir. Bu yöntemin en çok tercih edilmesinin nedenleri aşağıdaki gibi özetlenebilir;

- SRA parametrik olmayan bir modeldir bu bağlamda birçok istatistik tekniğine (lojistik regresyon, kümeleme analizi, çoklu regresyon, diskriminant analizi) alternatiftir.
- Bağımsız değişken türünde kısıtlama yoktur kesikli veya sürekli olabilir.
- Model ile ilgili herhangi bir varsayım gerekmediğinden değişkenlere dönüşüm (logaritma ya da karakök transformasyonları gibi) uygulanmasına gerek duyulmaz.

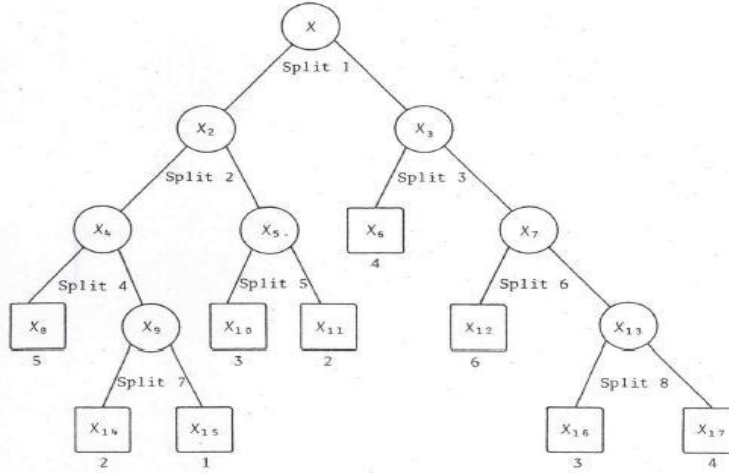
- Model ağaç şeklinde olduğundan bağımlı ve bağımsız değişkenler arasındaki ilişki kolaylıkla yorumlanabilir.
- SRA bağımlı değişken için olabilecek bütün bağımsız değişkenlere ve bunların bütün kombinasyonlarına baktığından model daha geniş bir bakış açısı sağlar.
- SRA oldukça karmaşık veri setlerinde bile doğru tahmin yapabilir ayrıca değişkenler için kayıp veya eksik gözlemlerin olmasından ve aykırı değerlerden etkilenmez.
- Eğer ihtiyaç duyulursa aynı bağımsız değişken aynı ağaçta farklı ayırma değerleriyle (cut off) kullanılabilir (Lewis, 2000).

SRA, parametrik yöntemleri etkileyen outlier (aşırı sapan gözlemler), lineerlik, değişen varyans veya dağılımsal hata şekli gibi durumlardan etkilenmez. Sapan değerler düğüm içinde izole edilirler ve bölünmeye etki etmezler. Parametrik modellerdeki durumların aksine SRA lineer değişkenleri vekil ayrıçlarda kullanır. (Baysal, 2011). SRA etkin olarak çok yönlülük ile ilgilenir. Analiz için çok sayıda değişken kullanıldığında sadece birkaç önemli değişken kullanarak faydalı sonuçlar üretir. SRA'nın önemli bir zayıf yönü parabolik modellere dayandırılmış olmasıdır. Hem ihtimal seviyeleri hem de güven sınırları yoktur. Yeni bir veri setini sınıflamak için kullanılan bir regresyon ağacından çıkarılan veya tahmin edilen tahminleyiciler ile ilgili bir önemlilik seviyesi ya da güven sınırları yoktur (Oruçoğlu, 2011). Bağımlı değişkenin türüne göre Sınıflama ve Regresyon Ağacı olarak ayrılan yöntem iki başlık altında detaylı olarak açıklanmıştır.

2.1 Sınıflama Ağaçları

Sınıflama çalışmalarının asıl amacı, doğru bir sınıflayıcı üretmek ve problemin altında yatan tahmin edici yapıyı ortaya çıkarmaktır. Sınıflama sürecindeki en önemli kriter sadece veri setine yönelik doğru bir sınıflayıcı üretmek değildir. Bir gözlemin hangi sınıfta yer alacağına belirlenmesinde; hangi değişkenlerin ya da değişkenler arası etkileşimin, olayı çözmeye faydalı olacağına anlaşılmasına çalışılması gerekmektedir (Breiman L. , Freidman, Olshen, & Stone, 1993).

Sınıflama ağacı 3 aşamadan meydana gelmektedir. Bunlar, "ağacın oluşturulması", "budama (pruning)" ve "en uygun ağaç yapısının seçimi" şeklindedir. Maksimum düzeyde homojen alt sınıflar (ağaçlar) oluşturma ilkesine dayalı çalışan sınıflama ve regresyon ağacı algoritması, "ağacın oluşturulması" kısmında olabilecek en fazla sayıda alt ağaçları belirlemektedir. Ancak, alt ağaçlar arasında bağımlı değişkenle önemli ilişkiler gösteren ağaçları seçmek gerekmektedir. Bundan dolayı algoritmanın ikinci kısmı olan "budama" modülü devreye girmektedir. Böylece budamadan sonra "en uygun ağaç yapısının seçimi" ile sınıflama ağacı elde edilebilmektedir.



Şekil 2 İkili Sınıflama Ağacı Alt Küme Örneği

Kaynak: (Breiman L. , Freidman, Olshen, & Stone, 1993)

Şekil 2.1’de görüldüğü üzere, $X = X_2 \cup X_3$ benzer şekilde,

$$X_4 \cup X_5 = X_2,$$

$X_6 \cup X_7 = X_3$ olarak her bir alt küme bir diğey adıyla düğüm, azalan şekilde iki alt küme ayrılmaktadır.

$X_6, X_8, X_{10}, X_{11}, X_{12}, X_{14}, X_{15}, X_{16}, X_{17}$ alt kümelerinin ise bölünme yapmadığı görülmektedir. Bu yüzden bu alt kümeler uç alt kümeler olarak adlandırılır. Uç alt kümeler bitiş noktası olduğu için her uç alt küme, altında numarasının gösterildiği bir sınıfa (C) atanır. Aynı sınıfa atanmış iki veya daha çok alt küme olabilir. Şekilde de görüldüğü gibi X_{10} ve X_{16} aynı sınıf olan 3. sınıfa atanmıştır.

$$X_2 = \{x; x_1 \leq c\}, \quad X_3 = \{x; x_1 > c\} \quad [2.1]$$

X bir uç kümeye atandığında onun sınıf tahmini, uç kümesinde yazan sınıf etiketi ile gösterilir. Bu noktada ağaç teorisindeki terminoloji yeniden şekillenmektedir ve anlatılmak istenen;

Bir t düğümü = X matrisinin bir alt kümesi ve

$$\text{Kök düğüm } t_1 = X$$

olur. Bu yüzden uç alt kümeler artık; uç düğümler olarak adlandırılır. Benzer şekilde uç olmayan alt kümeler; uç olmayan düğümler olarak adlandırılır (Breiman L. , Freidman, Olshen, & Stone, 1993).

Sınıflama ağaçlarında ayırmalar X bağımsız değişkeninin sürekli veya kategorik oluşuna göre iki şekilde tanımlanabilir. Eğer X sürekli bir değişken ise ayıraç “s (cut off değeri) reel bir sayı olmak üzere $x_n < s$ midir?” sorusudur. Deney üniteleri eğer soruya alınan yanıt evet ise sol düğüme, hayır ise sağ düğüme atanır. Söz konusu sürekli bağımsız değişkenin k adet farklı değeri var ise bu durumda en fazla k-1 adet

ayırac dolayısı ile k-1 adet farklı ayırma mümkündür.

Eğer X kategorik bir değişken ise ayırac “ $A \subset X$ olmak üzere, $x_n \in A$ mıdır?” sorusudur. Deneş ünıtelere eęer soruya alınan yanıt evet ise sol düęüme, hayır ise saę düęüme atanır. Söz konusu kategorik baęımsız deęişken k adet farklı kategori içeriyorsa bu durumda X’in en fazla 2^{k-1} adet (boş küme hariç) alt kümesi vardır ve dolayısı ile $2^{k-1} - 1$ adet farklı ayırma mümkündür (De’ath & Fabricius, 2008).

Sınıflama aęaçlarında herhangi bir t düęümünde mümkün olan tüm olası ayırmalar belirlendikten sonra, her bir olası ayırma için ayırmanın uygunluk derecesi hesaplanır. Ayırmanın uygunluk derecesi, ayırma fonksiyonu yardımıyla hesaplandıktan sonra, maksimum uygunluk derecesine sahip ayırma en iyi ayırma olarak seçilir.

Ayırmalarda herhangi bir düęümün heterojenlik deęeri safsızlık ölçüsü olarak adlandırılır ve bu deęer safsızlık fonksiyonu kullanılarak hesaplanır. Safsızlık ölçüsü sıfır deęerini alıyorsa düęüm tamamen homojendir.

Safsızlık fonksiyonu tahmin etmek için bir düęümdeki (t), j sınıfının oranını bilmek gerekir. (Sezgin, 2006). Bunun için; $\pi(j)$ yani j sınıfının önsel olasılığı eřitlik [2.2]’deki gibi hesaplanır.

$$\pi(j) = \frac{N_j}{N} \quad [2.2]$$

Böylelikle her düęüme ait ilgili sınıfa (j) düşen gözlem sayısının, tüm gözlem sayısına bölünmesi ile gözlemin j sınıfında olma olasılığı yani j sınıfının önsel olasılığı bulunmuş olur.

Hesaplanan $\pi(j)$ ařaęıdaki formüle yerleřtirilirse; eřitlik [2.3] bulunmuş olunur.

$$p(j, t) = \pi(j) \frac{N_j(t)}{N_j} \quad [2.3]$$

Burada;

$N_j(t)$: j sınıfına ait t düęümünün gözlem sayısıdır.

$p(j, t)$: hem j sınıfı hem de t düęümünde bulunan bir gözlemin olasılığıdır.

$$p(j, t) = \frac{N_j}{N} \frac{N_j(t)}{N_j} \rightarrow p(j, t) = \frac{N_j(t)}{N} \quad [2.4]$$

deęeri bulunmuş olur. Böylelikle $p(j, t)$ formülasyonu ile t düęümünde bulunan ve j sınıfına ait olan bir gözlemin olasılığı hesaplanmış olur.

Ayrıca; t düğümüne ait olan herhangi bir gözlemin olasılığı eşitlik [2.5]'teki gibi bulunur

$$p(t) = \sum_j p(j, t) \quad [2.5]$$

$$p\left(\frac{j}{t}\right) = \frac{p(j,t)}{p(t)} = \frac{N_j(t)}{N(t)} \quad [2.6]$$

$p(j/t)$: t düğümündeki j sınıfında olmanın koşullu olasılığıdır. Yani t düğümüne düşmüş olduğu bilinen bir gözlemin j sınıfında olması olasılığı olup bir koşullu olasılıktır. $\sum_j p\left(\frac{j}{t}\right) = 1$ olduğu bilinmektedir (Saitoğlu, 2015).

Safsızlık fonksiyonundaki değişimin maksimum olması ile en iyi bölünme kuralının seçilmesi sağlanır (Sezgin, 2006). Buna göre, belli bir (δ) bölünme değeri için bir t düğümünün safsızlık fonksiyonundaki değişim yani safsızlıktaki azalış eşitlik [2.7]'daki gibi tanımlanır:

$$\Delta i(\Delta\delta, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad [2.7]$$

p_L : sol düğümdeki (t_L) gözlemlerin oranı (olasılığı), yani t ana düğümden sol tarafa ayrılan gözlemler.

p_R : sağ düğümdeki (t_R) gözlemlerin oranı (olasılığı), yani t ana düğümden sağ tarafa ayrılan gözlemler.

Böylelikle en iyi bölme kriteri olan $\phi(\delta, t)$ yani bir diğer adıyla safsızlık fonksiyonu, $\Delta i(\delta, t)$ 'ye yani bir δ bölünme değeri için bir t düğümünün safsızlık fonksiyonundaki değişim ifadesine denk gelmektedir. $\phi(\delta, t)$ değerini maksimize eden δ^* değeri için t düğümünün bölünmesi gerçekleştirilir.

Sınıflama ağaçlarında kullanılan birçok safsızlık ölçüsü (Gini Diversity Index, Twoing, Chi-square, G-square) vardır. Safsızlık ölçüsü herhangi bir t düğümü için en iyi ayırmanın seçimini önemli bir şekilde etkilemektedir. Bu sebepten safsızlık ölçüleri literatürde en iyi ayırma kriteri olarak da bilinirler. En yaygın kullanılan ayırma kriterleri Gini Diversity Index (Gini) ve Twoing Kuralıdır.

2.1.1 Gini Ayırma Kriteri

Safsızlık fonksiyonu kriterlerinden Gini algoritmasındaki amaç her adımda en büyük veri kümesini elde etmektir. Böylece en iyi bölme durumu elde edilmiş olunacaktır. Ayrıca bölme işleminden sonra ilgilenmediğimiz kısım izole edilmiş olur (Bozan, 2010).

Bu kriter değişken değerlerinin sol ve sağda olmak üzere iki bölüme ayrılması esasına dayanır (Özkan Y. , Veri Madenciliği Yöntemleri, 2016). Düğümlerin ikiye ayrılabilmesi için bölünme noktalarının belirlenmesi gerekir. En iyi bölünmeyi seçmek için geliştirilen gini kriteri ya da bir diğer adıyla gini indeksi bir safsızlık ölçütü (fonksiyonu) olarak da anlatılmaktadır. Daha önce belirtildiği gibi; $p(j/t)$ ve $j=1, \dots, J$, t düğümündeki j sınıfının olasılığı olmak üzere, t düğümünün safsızlığının ölçütü

$i(t) = \phi(p(1/t), \dots, p(J/t))$ formülasyonu ile tanımlanır. Düğümün ve doğal olarak ağacın safsızlığını düşüren bölünme araştırılır. Böylelikle gini kriterine göre oluşturulan safsızlık fonksiyonu eşitlik [2.8]'deki gibi tanımlanır.

$$i(t) = \sum_{j \neq i} p\left(\frac{j}{t}\right) p\left(\frac{i}{t}\right) \quad [2.8]$$

Bu formül ile t düğümüne ait gini indeksi bulunur. Bu formül t düğümünün sol ve sağ düğümüne atanan değişkenin kategori değerleri için ayrı ayrı hesaplanır (Saitoğlu, 2015).

Daha önce de bahsedildiği gibi en iyi bölme kuralı safsızlık fonksiyonundaki değişimi maksimize eden gini değerine göre belirlenir. Maksimum gini değişimini veren değişkenin ilgili kategorisine göre bölünme yapılır (Sezgin, 2006).

2.1.2 Twoing Ayırma Kriteri

Twoing algoritması ise, Gini' ye göre daha dengeli bir yapı sunar. Bunun nedeni ise her defasında ana ve yavru düğümlerin %50' sini içermeye çalışmasıdır. Bundan dolayı Gini' ye göre bir yavaşlama olacaktır (Bozan, 2010). Twoing ilk olarak, tek bir sınıfı diğerinden ayırmak yerine, düğüme ait verilerin %50'sini kapsayan ve birbirine benzemeyen sınıfları ayırmaya çalışır.

Daha önce bahsedildiği üzere çok sınıflı problemlerde gözlemlerin J adet sınıfa atandığını varsaydığımızda sınıf sayısı 1,2,...J adet olmak üzere, C sınıf seti olarak adlandırılır ve $C = \{1,2,\dots,J\}$ ile gösterilir. Her düğümde sınıflar sağ ve sol olmak üzere iki üst sınıfa (superclasses) atanır. Gözlemlerin $\{g=1,2,\dots,k,\dots,n\}$ bir kısmının atandığı sınıf C_1 , geri kalanının atandığı sınıf C_2 olup;

$$C_1 = \{1,2,\dots,k\}, \quad C_2 = C - C_1$$

Yani $C_2 = \{n-1, n-2, \dots, n-k\}$ 'dir. Buna göre, $\Delta i(\delta, t)$ yani bir δ bölünme değeri için bir t düğümünün safsızlık fonksiyonundaki değişim C_1 'in seçimine bağlı olur ve notasyon $\Delta i(\delta, t; C_1)$ ile ifade edilir. Problem artık $\Delta i(\delta, t; C_1)$ değerini maksimize eden $\delta^*(C_1)$ bölünmesini bulmaya dönüşür ve ilgili C_1^* bulunur. Bu durumda notasyon; $\Delta i(\delta^*(C_1), t; C_1)$ ile ifade edilir ve bölünme $\delta^*(C_1)$ durumunu sağlayan düğüm için gerçekleştirilir. Bu mantık, sınıfları ikiye ayırmak için aynı şekilde uygulanır.

Buna göre twoing bölme kuralının teoremini şu şekilde açıklayabiliriz; iki sınıf kriteri $p(1/t) p(2/t)$ altında bir δ bölünme değeri ve bir üst sınıf $C_1(\delta)$ değeri için $\Delta i(\delta^*(C_1), t; C_1)$ ifadesi maksimize edilir. Burada $C_1(\delta)$:

$$C_1(\delta) = \{j: p(j/t_L) \geq p(j/t_R)\} \quad [2.9]$$

ile ifade edilir. Buna göre twoing bölme kuralına göre hesaplanan safsızlık değişimi eşitlik [2.10] 'da ki gibi formüle edilir (Saitoğlu, 2015).

$$\max_{C_i} \Delta i(\delta, t, C_1) = \frac{P_L P_R}{4} \left[\sum_{j=1}^J p\left(\frac{j}{t_L}\right) p\left(\frac{j}{t_R}\right) \right]^2 \quad [2.10]$$

2.1.3 Gini ve Twoing Ayırma Kriterlerinin Karşılaştırılması

Düğümdeki mümkün olan en iyi ayırmayı gerçekleştirmek için ayırma kuralı seçilirken aşağıdaki faktörler göz önünde bulundurulmalıdır.

- Kategorik bağımlı değişkenin seviye sayısı iki ise ve analizin 0.50'den daha az hata oranına sahip olacağı tahmin ediliyorsa ayırma kuralı olarak Gini kuralı tercih edilir.
- Kategorik bağımlı değişkenin seviye sayısı iki ise ve analizin 0.80'den daha az hata oranına sahip olacağı tahmin ediliyorsa ayırma kuralı olarak Twoing kuralı tercih edilmelidir.
- Bağımlı değişkenin seviye sayısı 4'ten daha büyük olduğu koşullarda Twoing kuralı Gini'den daha doğru bir seçimdir (Temel, 2004).

2.1.4 Sınıflama Ağaçlarında Maksimum Boyuttaki Ağacın Oluşturulması ve Ağacının Büyüme Sürecinin Durdurulması

Başlangıç veri kümesinin en son gözleme kadar bölünmesi ile maksimum ağaç oluşturulur. Yani her bir gözlem uç düğümlerde tek bir sınıfta yer alınca maksimum boyuttaki ağaç oluşmuş olur (Timofeev, 2004).

Kök düğüm t değeri için aday bölünme ayırma (δ) değeri vardır. Her bölünmede, her gözlemin x_i değerinin $x_i \leq \delta$ olup olmadığına bakılır ve sonucuna göre sol yada sağ düğüme ataması yapılır. Daha sonrasında her bölünme için bölme uyum kriterin (gini yada twoing) tahmin edilir ve bu kritere göre en iyi bölünme seçilir. Bu aşamalar her bağımsız değişken için ardışık olarak tekrar edilir. Bu süreç tekrarlı bölünme (recursive partitioning) olarak adlandırılır.

Sınıflama ağacı yönteminde ikili bağımlı değişkenin saflığına karar verilirken Gini katışıklık ölçümü kullanılmaktadır.

Bir t düğümü için, Gini katışıklık indeksi ($g(t)$) aşağıdaki formülle belirlenmektedir.

$$g(t) = \sum_{j \neq i} p\left(\frac{j}{t}\right) p\left(\frac{i}{t}\right) \quad [2.11]$$

Burada i ve j hedef (bağımlı) değişkenin kategorileridir. Türlerin dağılımına yönelik modellemelerde bağımlı değişkenlerin ikili (var-yok) kategorilerinden oluştuğundan dolayı indeks için eşitlik aşağıdaki gibidir.

$$g(t) = 2p\left(\frac{1}{t}\right) p\left(\frac{2}{t}\right) \quad [2.12]$$

Düğümdeki bütün yenilenmiş kodlar sadece bir kategoriye ait olduğunda (bu durumda düğüm saftır indeks değeri sifıra eşittir) bir düğümün en iyi kestirimini yapacak bağımsız değişkeni bulmak için bağımsız değişken setindeki her değişken değerlendirilir ve en iyi değere sahip olan değişken seçilir. Herhangi bir t düğümü için düğümün bir aday ayırıcısı olan s hem sağ taraf ayırımını (t_r) hem de sol taraf ayırımını (t_l) gerçekleştirir.

Değer eşitlik [2.13] ' teki formül ile belirlenir.

$$\emptyset(s,t)= g(t) - p_l g(t_l) - p_r g(t_r) \quad [2.13]$$

Burada p_r sağ taraftaki bağımlı t düğümündeki durumların oranını verirken p_l sol taraftaki t düğümündeki durumların oranını vermektedir. Her bir düğümde ikili s 'in bir aday S seti belirlenebilmektedir ve kök düğümü olan t_1 den başlayarak ayırıcı s^* bütün muhtemel S 'ler arasında daha büyük bir katışıklık azalama değeri ile aranmaktadır.

$$\emptyset(s^*,t) = \max_{s \in S} \phi(s,t) \quad [2.14]$$

İdeal bir ayırıcı s veri setini $g(t_l)=g(t_r)=0$ olacak şekilde iki alt gruba ayırmaktadır. Tekrarlamalı bölme algoritması, homojenlik sağlanana ve bazen tek bir durum bir sınıfta kalana kadar devam etmektedir. Bütün nihayi düğümlerin en ideal saflığa gelmesi ile maksimum ağaç elde edilmiş olur. Maksimum boyuttaki ağacın sonlandırılması ise, her düğümün uç düğüm olup olmaması ile neticelenir ve her uç düğüme bir sınıf atanınca sonuçlanır (Özkan K. , 2012).

Sonlandırma kuralı basit bir süreci içinde barındırmaktadır. Bir düğümün sonlandırılabilmesi için ağaçtaki bölünmenin bitmiş olması gerekir. Bölme süreci ağacın safsızlığındaki azalış artık imkansılaşmaya başlayınca son bulmaktadır. Buna göre ağacın safsızlığındaki değişim $I(T)$ olmak üzere;

$$\Delta I(T)=\Delta i(t)p(t) \quad [2.15]$$

formülasyonu ile belirlenmektedir. Ağacın safsızlığındaki azalış son bulunca ağaç sonlanır ve uç düğümler belirlenmiş olur. Ağaç çiziminde daha önce de belirtildiği gibi uç düğümler dikdörtgen şeklinde, diğer düğümler ise daire şeklinde olur (Sezgin, 2006).

Bu tanımlardan sonra, bir sınıflama ağacının oluşturulması için gerekli adımlar aşağıdaki şekilde özetlenebilir;

1. Adım: Düğümde yer alan deney ünitelerinin içerdiği bağımsız değişkenler bu değişkenlerin birbirleri ile kombinasyonlarının tanımlı bulunduğu aralıklardaki tüm olası değerleri birer ayıraç olarak varsayıp, mümkün olan tüm olası ayırmaların belirlenmesi,
2. Adım: Mümkün olan her bir ayırma için o ayırmanın uygunluk derecesini, ayırma fonksiyonu yardımıyla hesaplayarak maksimum uygunluk derecesine sahip ayırmanın belirlenmesi,
3. Adım: En iyi ayırmayı yapan ayırıcının, t düğümüne uygulanması ve ortaya çıkacak sol ve sağ çocuk düğümlerinin her birine en uygun sınıfın tahmin edilmesidir.

Yukarıda sıralanan adımlar kök düğümünden başlayarak, daha sonra ortaya çıkacak her düğüm için tekrarlanır. Sınıflama ağacı, her bir düğüm noktası bu şekilde ikiye

ayrılarak büyür ve bu büyüme; her düğümde grup içi homojenlik söz konusu ise, ağacın düzey sayısından analizi yürüten kişi tarafından bir sınırlama yapıldıysa ve yeni oluşacak düğümlerde fazla bir değişiklik yaratmıyorsa durur (Temel, 2004).

2.2 Regresyon Ağaçları

Bağımlı değişkenin sürekli olduğu durumlarda SRA regresyon ağacı olarak adlandırılır. Regresyon ağacı yöntemi, bağımsız değişkenlerin dağılımına ilişkin herhangi bir varsayım gerektirmemesi, çoklu bağlılık (Multicollinearity), sapan değerler (outliers) ve kayıp gözlemlerden (missing observation) etkilenmemesi gibi avantajlardan dolayı tercih edilmektedir (Akkartal & Mendeş, 2009). Regresyon ağacı yöntemi, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ve etkileşimi tespit eden, popülasyonu homojen alt sınıflara ayıran parametrik olmayan bir analiz yöntemidir (Kayri & Boysan, 2008).

Regresyon ağaçlarını oluşturma adımları aşağıda belirtildiği gibidir:

1. Adım: Regresyon analizine başlarken regresyon ağacında kök düğüm ile başlanır.
2. Adım: m_c herhangi bir "c" dalına ilişkin tahmin değerini, V_c ; dalına ayırma varyansını ve n_c 'de bu daldaki gözlem sayısını belirtmek üzere ayırma işlemi $S = \sum \sum (Y_i - m_c)^2$ biçiminde ifade edilen S değeri minimum oluncaya kadar devam ettirilir. Bu ifade $S = \sum n_c V_c$ biçiminde de gösterilebilir, m_c ise $\sum c Y_i / n_c$ dir.
3. Adım: Düğüm içerisinde bütün bireyler dikkate alınan tüm bağımsız değişkenler için aynı değeri alıyorsa ağacın bölünme işlemi durdurulur. Aksi takdirde S minimize edilebilecek en son noktaya kadar düşürülür. Eğer S'deki en büyük azalma belirlenen δ gibi bir eşik değerinden küçük ise bölünme sonlanır.
4. Adım: Her bir yeni düğüm için 1. Adıma geri dönülür (Küçüköğlü, 2010).

Regresyon Ağacının sahip olduğu algoritma, benzerlik gösteren değişkenlerin aynı ağaç düğümünde toplanmasına dayalı olup, bütün oluşturduğu alt dalları (sub-branches) bağımlı değişken olan kök düğüme bağlamayla son bulmaktadır (Breiman L. , Freidman, Olshen, & Stone, 1984). Regresyon Ağacına ait bu algoritma temelde 3 unsurdan meydana gelmektedir. Bunlar, "ağacın oluşturulması", "budama (pruning)" ve "en uygun ağaç yapısının seçimi" şeklindedir. Maksimum düzeyde homojen alt sınıflar (ağaçlar) oluşturma ilkesine dayalı çalışan sınıflama ve regresyon ağacı algoritması, "ağacın oluşturulması" kısmında olabilecek en fazla sayıda alt ağaçları belirlemektedir. Ancak alt ağaçlar arasında bağımlı değişkenle önemli ilişkiler gösteren ağaçları seçmek gerekmektedir. Bundan dolayı algoritmanın ikinci kısmı olan "budama" modülü devreye girmektedir (Baysal, 2011).

Regresyon ağacı analizinin budama sürecinde yüksek hatalı sınıflamaya neden olan diğer bir ifade ile modelin tahmin edici değerine az katkıda bulunan düğümler ya da dallar atılmaktadır. Budama işlemi, cost-complexity parametresine göre yapılmakta, bu parametre, değer ya da maliyet fonksiyonu, hatalı sınıflama ve ağaç büyüklüğü (regresyon ağacındaki düğüm sayısı) değerleri dikkate alınarak hesaplanmaktadır (Akkartal & Mendeş, 2009). Budama işlemi, ağaçta oluşmuş fakat sonucu

etkilemeyen ve sınıflanmaya herhangi bir katkısı olmayan dalların ağaçtan alınmasıyla bir bakıma ağacın gereksiz ayrıntılardan arındırılmasıdır. Budama uygulanmasının amacı önceki aşamalarda modele dahil edilmiş değişkenlerin ileriki aşamalarda tekrar modele dahil olmasını engellemektir. Budama süreci, en az katkı sağlayan düğümden başlayarak, önemli katkı sağlayan düğümler kalıncaya kadar devam ettirilir. Budanmış ağaç diğer yarısına gelene kadar tekrar budanır. Bu budama işlemi ağacın boyu artık değişmeye kadar devam eder. Böylece budamadan sonra “en uygun ağaç yapısının seçimi” ile regresyon ağacı elde edilebilmektedir. Regresyondaki ağaç oluşum süreci, sınıflama ağacındakinden daha basittir ve ağacın büyümesi için kullanılan safsızlık kriteri, aynı zamanda ağacın budanmasında da kullanılmaktadır. Bunun yanısıra regresyon ağacının oluşumu için herhangi bir sınırlama yoktur. Her gözlem eşit ağırlığa sahiptir. Regresyon ağaçlarında kullanılan safsızlık kriteri en küçük kareli sapma hesabına dayanmaktadır (Küçüköğlü, 2010).

Bu süreçteki ağaç oluşum yapısı bazı farklılıklarına rağmen sınıflama ağacı ile benzerlik gösterir. Sınıflama ağaçlarında bağımsız değişkenler, uç düğüm olarak sonlanan bir sınıfa atanmaya çalışır. Oysa regresyon ağaçlarında, bağımlı değişken sürekli olduğu için her uç düğüme atanan değer, bağımlı değişkenin sayısal bir tahminidir. Bu sebepten ötürü regresyon ağaçlarındaki bölme kriteri de farklılık göstermektedir (Sezgin, 2006). Regresyon ağaçlarında sınıflar yoktur. Bu sebepten regresyon ağacı tekniğinde sınıflama ayırım kuralları Gini indeksi kullanılarak uygulanamaz. Regresyon ağacındaki ayırımlar iki sonuçlanan düğüm için tahmin edilen toplam varyansın minimize olmasının gerekliliği anlamına gelen “artıkların karelerini azaltma algoritmasına” göre gerçekleştirilir (Özkan K. , 2012). Regresyon ağacında en sık kullanılan bölme uyum kriterleri en küçük kareli sapma (Least Squared Deviation: LSD) yöntemidir.

2.2.1 En Küçük Kareli Sapma (LSD)

Sınıflama ve Regresyon Ağaçlarında kullanılan bölme uyum kriteri bağımlı değişken sürekli olduğunda yani regresyon ağacı problemlerinde kullanılan ve safsızlık fonksiyonu dayalı bölme kriteri olan en küçük kareli sapma yöntemidir. En küçük kareli sapma yöntemi ile en küçük kare hata (least squares error) kriteri minimize edilerek parametreler tahmin edilir. Regresyon ağaçlarında başlangıç veri kümesi aracılığıyla bir kestirimci $d(x)$ ve onun hata tahmini olarak ifade edilen $R^*(d)$ değeri tahmin edilir. R^* tahmin etmek için bir çok yol bulunmaktadır. Yerine koyma (resubstitution) en yaygın olan tahmin yöntemidir. Diğer tahmin yöntemi ise ağırlıklandırılmış varyans yöntemidir. Bunun için ilk olarak yerine koyma tahmin yöntemi anlatılmıştır.

Yerine koyma hatası eşitlik [2.16] da ki gibidir:

$$R(d)=\frac{1}{N}\sum_{n=1}^N(y_n - d(x_n))^2 \quad [2.16]$$

Bu eşitlikten anlaşıldığı üzere en iyi bölünme değeri, düğüm içi kareler toplamının ortalamasını veya yerine koyma tahminini minimize eden değerdir (Saitoğlu, 2015).

Belirtildiği üzere yerine koyma tahminin alternatif yolu ağırlıklandırılmış varyansı kullanmaktır. Buna göre, $p(t) = N(t)/N$ ifadesi t düğümüne düşen raslantısal olarak seçilmiş bir gözlemin olasılığı için yerine koyma tahmini olsun. Buna göre t düğümünün varyansı eşitlik [2.17] de ki gibidir.

$$S^2(t) = \frac{1}{N(t)} \sum_{x_n \in t} (y_n - \bar{y}(t))^2 \quad [2.17]$$

Böylece $R(t) = s^2(t)p(t)$ ve buradan;

$$R(t) = \sum_{t \in \hat{T}} s^2(t)p(t) \quad [2.18]$$

$S^2(t)$, t düğümündeki y_n değerlerinin örnek varyansıdır. Böylece t düğümüne ait en iyi bölünmenin ağırlıklı varyansı minimize etmesi beklenir (Saitoğlu, 2015).

Sürekli hedef değişkenleri için en küçük kareli sapma heterojenlik ölçüsü kullanılmaktadır. LSD ölçüsü $R(t)$, t düğümü için basit (ağırlıklandırılmış) düğüm içi varyansıdır ve düğüm için risk tahminine eşittir. $R(t)$ 'n formülü aşağıdaki şekildedir:

$$R(t) = \frac{1}{N_w(t)} \sum_{i \in t} W_n f_n (y_i - \bar{y}(t))^2 \quad [2.19]$$

$N_w(t)$, t düğümündeki ağırlıklandırılmış düğüm sayısı, W_n i düğümü için mevcut ise ağırlıklandırılma değişken değeri, f_n mevcut ise frekans değişkeninin değerini, y_i hedef değişkenin değerini ve $\bar{y}(t)$ ise t düğümü için ağırlıklı ortalamayı göstermektedir (Oğuzlar, CART Analizi ile Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi, 2004).

2.2.2 Regresyon Ağaçlarında Maksimum Boyuttaki Ağacın Oluşturulması ve Büyüme Sürecinin Durdurulması

Ağaç oluşum sürecinde oluşturulan en büyük boyuttaki ağaç maksimum boyuttaki ağaçtır. Bu bağlamda regresyon ağaçlarındaki maksimum boyuttaki ağaç oluşumu sınıflama ağacının maksimum boyuttaki ağaç oluşumu ile benzer şekilde oluşturulmaktadır. Başlangıçta veri kümesinin bölünmesi için regresyon ağaçlarında kullanılan bölme uyum kriteri safsızlık kriteri en küçük kareli sapma hesabı ile belirlenmeye çalışılır ve sınıflama ağacında olduğu gibi ikili bölünmeler gerçekleştirilir. Regresyon ağacı yönteminde her düğümde minimizasyon (azaltma) problemi aşağıda gibi çözülür.

$$\arg \min [P_1 \text{Var}(Y_1) + P_R \text{Var}(Y_R)], \quad x_j \leq x_j^R, \quad j=1, \dots, M \quad [2.20]$$

Burada P_1 ve P_R sırası ile sol ve sağ düğümlerin olasılıklarıdır. M eğitim setindeki değişkenlerin sayılarıdır. Değişken j " x_j " olarak gösterilmektedir. x_j^R ise değişken x_j nin en iyi ayırım değerini göstermektedir. $\text{Var}(Y_1)$, $\text{Var}(Y_R)$ karşılıklı sağ ve sol alt düğümler için sorumlu vektörlerdir.

$x_j \leq x_j^R, j=1, \dots, M$ optimal ayırım sorgulaması anlamına gelmektedir.

Artıkların karelerini azaltma algoritması Gini ayırım kurallarına benzemektedir. Eğer sınıf k 'nin nesnelere değeri "1", diğer sınıfların nesnelere değeri "0" ataması yapılırsa, zamanla bu değerlerin örnek varyansı $p(k/t)(1-p(k/t))$ 'e eşit olur. Katışıklık ölçümü $i(t)$ eşitlik [2.21] de ki gibi bulunur.

$$i(t) = 1 - \sum_{k=1}^K p^2(k/t) \quad [2.21]$$

Burada $p(k/t)$ düğüm t içinde sınıf k 'nin koşullara bağlı özelliklerini, K sınıf sayısını, k sınıf indeksini ve t düğüm indeksini göstermektedir. Maksimum boyutta ağaca ulaşmış regresyon ağacının büyüme sürecinin durdurulması ise sınıflama ağacında olduğu gibi uç düğüm kuralına göre sonlandırılmaktadır. Sınıflama ağaçlarından farkı ise regresyon ağaçlarında bir düğüm aşağıdaki koşulları sağladığında sonlandırılmaktadır.

$$N(t) < N(\min)$$

$N(\min)$ genellikle 5 olarak alınır (Saitoğlu, 2015).

2.3 Sınıflama ve Regresyon Ağaçlarında Optimum Ağacın Belirlenmesi

Modelin oluşturulmasından sonraki en büyük problem oluşturulan modelin doğruluğudur. Sınıflandırma ve Regresyon Ağacı tekniği ile oluşturulmuş ağaç modellerinde ağacın karmaşıklığı ve doğruluk derecesi temel problemdir. Bu problemi ortadan kaldırmak için optimum boyuttaki ağacın belirlenmesi gerekir. Optimum boyuttaki ağacın belirlenmesinde aşağıda anlatılan model seçim kriterleri kullanılır. Optimal ağacın seçimi için kullanılan *maliyet-karmaşıklık-budama yöntemi* (*cost-complexity pruning method*) hatalı sınıflama oranı ile ağacın karmaşıklığı arasındaki dengeyi sağlar ve bu işlem matematiksel olarak;

$$R_a^{(T)} = R(T) + \alpha \cdot T \text{ şeklinde ifade edilir. Burada;}$$

$$R_a^{(T)} = \text{Maliyet-karmaşıklık ölçüsünü,}$$

$$R(T) = T \text{ ağacı için hesaplanan hata oranını,}$$

$$T = T \text{ ağacındaki terminal düğüm sayısını,}$$

α = Ağaçtaki her terminal düğüm için belirlenen ceza katsayısını ($\alpha \geq 0$) gösterir (Temel, 2004).

Maliyet-karmaşıklık-budama yöntemiüne göre maksimal ağaç, maliyet-karmaşıklık ölçüsü minimum değerine ulaşmaya kadar budanır ve optimum ağaç elde edilir. Optimum ağacın belirlenmesi için kullanılan ölçü sınıflama ağaçlarında ve regresyon ağaçlarında farklılık göstermektedir;

Sınıflama Ağaçlarında, en küçük maliyet karmaşıklık ölçüsü (minimal cost complexity measure) optimum boyuttaki ağacı belirlemek için kullanılan etkili bir yöntemdir. Ağacı bir bütün olarak inceleyip, ağacın doğruluğuna önemli derecede katkısı olmayan alt dalların yani zayıf bölünmelerin budanması işlemi için kullanılan bir

ölçüttür. Hem yanlış sınıflama oranı hem de ağacın karmaşıklığı aynı anda minimize edilmeye çalışılır (IBM SPSS Modeler 15 Algorithms Guide). Maliyet-karmaşıklık ölçüsünde α değerinin artması optimal ağaçta daha az terminal düğümünün yer almasına yol açar. Bir başka ifade ile α değeri arttıkça budama artar (Temel, 2004).

Regresyon Ağaçlarında, optimum boyuttaki ağaca ulaşmak için sınıflama ağaçlarında kullanılan maliyet-karmaşıklık ölçüsü yerine, bir hata-karmaşıklık (error-complexity) ölçüsü kullanılır. Böylelikle regresyon ağacının doğruluğu hata kareler ortalamasının tahmini ile belirlenir. Sınıflama ağaçlarında olduğu gibi her α için alt diziler, $T_{\max}, T_1, \dots, t_0$ tanımlanır ve her alt dizi için hata-karmaşıklık ölçüsü tahmin edilir. Böylelikle bu ölçüyü minimize eden ağaç optimum boyuttaki ağaç olarak kullanılır (Sezgin, 2006).



3 YAŞAM MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN BELİRLENMESİ

Bu bölümde önce yaşam memnuniyeti ve mutluluk kavramlarının anlamı ve önemi üzerinde durulmuş ve aralarındaki ilişki kısaca açıklanmıştır. Çalışmanın uygulama aşamasında ise yaşam memnuniyetini etkileyen faktörler sınıflama ağacı ve regresyon ağacı ile belirlenmiş ve sonuçlar yorumlanmıştır.

3.1 Yaşam Memnuniyeti ve Mutluluk Kavramı

Tartışmasız ki insanların yaşamları boyunca ortak istekleri mutluluktur, insanlar mutlu olmak için yaşarlar. Bu da hayatın amacı olan mutluluğun ne kadar önemli bir kavram olduğunu göstermektedir.

Mutluluk;

- Türk Dil Kurumu'nun Güncel Türkçe Sözlüğü'ne göre; bütün özelemlerle eksiksiz ve sürekli olarak ulaşılmaktan duyulan kıvanç durumu şeklinde tanımlanmaktadır;
- Felsefi Terimler Sözlüğü'ne göre, genellikle insanların kendilerine en yüksek amaç olarak koydukları değer ve bilinci dolduran tam bir doyguluk durumu;
- Çeşitli yazarlara göre, sevinçli anlardan oluşmakta, zaman ve ölüm kavramlarının unutulduğu an;
- Kişisel gelişimcilere göre; genellikle sosyal kazanımlara bağlı,
- Bir davranış bilim uzmanına göre ise neşe ve haz anları, tatmin duygusu ve iyi yaşamdır. (Giray, 2011)

Yüzyıllar boyu mutluluk kavramı üzerine düşünülmüş ve mutluluk kavramı sorgulanmış, insanlar mutlu olmanın yollarını aramışlardır. Mutluluk, M.Ö 300'lü yıllardan bu yana araştırılmaktadır. Felsefi açıdan bakıldığında; mutluluğun, ahlak bağlamında - iyi insan - kötü insan ayırımı ile ilişkilendirildiği görülmektedir (Babaoğlu, 2008).

Tarihi bu kadar eskiye dayanan mutluluk konusu, günümüzde de güncelliğini korumakta ve bu konuda çalışmalar yapılmaktadır. Dünyada yoğun bir şekilde tartışılmakta olan mutluluk kavramı, son dönemlerde tıp, psikoloji ve sosyoloji alanlarının yanısıra iktisat biliminin de ilgi alanına girmiştir. Mutluluğun, demografik, ekonomik, fiziksel çevre, sosyal çevre ve içinde yaşanan ülkenin sosyo-ekonomik durumu gibi değişkenlerin bileşimi sonucunda oluştuğunu belirtenlere göre, mutluluk ile yaşamdan alınan memnuniyet arasında büyük bir yakınlık söz konusudur (Şeker, 2009).

Yaşam memnuniyeti, bireyin kendi yaşam kalitesini bütünüyle değerlendirmesi sonucunda ulaştığı olumluluk derecesi anlamına gelmektedir. Aynı zamanda yaşam doyumunu olarak da ifade edilebilen memnuniyet ve haz içeren yaşam memnuniyeti, bireyin sürdürmekte olduğu yaşamından ne kadar hoşlandığını ifade etmektedir (Güler & Emeç, 2006). Yaşam memnuniyetinin başlıca belirleyicileri yaş, sağlık, stres,

yaşam şekli ve kişilik özellikleri olarak sıralanabilir (Chow, 2005). Öte yandan sosyal ilişkiler, başarı, doğa ile ilgili uğraşlar, cinsel faaliyet, beslenme türü, kitap okuma ve müzik dinleme gibi aktiviteler de yaşam memnuniyetini etkileyen faktörlerdendir (Şeker, 2009). Çoğu insan için mutluluk, yaşamdaki en önemli amaçtır (Gilman, Huebner, & Laughlin, 2000). Mutluluk, kişinin yaşamını kendi değerlerine göre dolu, anlamlı ve huzurlu bir şekilde algılama sürecidir.

Tanımlar incelendiğinde mutluluk ve yaşam memnuniyeti kavramlarının iç içe geçmiş kavramlar olduğu görülmektedir. Yaşam memnuniyetine dair ilk akademik çalışmalar 1925 yılında Fluegel tarafından yapılmaya başlanmıştır. Sübjektif refah düzeyinin ölçümüne dair modern çalışmaların öncüsü kabul edilen Fluegel, kişilerin günlük yaşamlarında karşılaştıkları olaylara verdikleri duygusal tepkileri derleyerek ruhsal yaşamlara ilişkin değerlendirmelerde bulunmuş, 2. Dünya Savaşı sonrasında yaşam memnuniyeti ile kişilerin mutluluğu arasındaki ilişkiler ağını araştırmıştır (Korkmaz, Yücel, Gürkan, & Germir, 2015). ABD'deki "Genel Sosyal Araştırması" ve Avrupa Birliği'nin "Euro-barometre" araştırmaları bu alanda en fazla bilinen araştırmalar olmasına rağmen literatürde, bir çok bilim adamı ve araştırma merkezi tarafından mutluluk üzerine yapılan araştırmalara rastlanmaktadır. 1970'li yıllara kadar bilimsel anlamda üzerinde çok fazla durulmayan mutluluk kavramı, özellikle 1974 yılında sadece "mutluluk" çalışmalarının yayınlandığı uluslararası bir derginin (Journal of Happiness Studies) yayın hayatına geçmesiyle üzerinde yoğun araştırmaların sürdürülmeye başlandığı bir kavram olmuştur (Şeker, 2009). Türkiyede yaşam memnuniyeti ile ilgili yapılan çalışmalara ait literatür incelendiğinde; yaşam memnuniyeti ya da mutluluk çalışmalarının çoğu spesifik olarak belirli sosyodemografik özelliklere sahip gruplar ya da bir şehir ile ilgilendiğinden Türkiye genelini temsil etmemektedir.

Türkiye'de ilk Yaşam Memnuniyeti Araştırması; TÜİK tarafından 2003'te Hanehalkı Bütçe Anketi'ne bir modül olarak eklenmiş ve uygulanmıştır. 2004 Nisan ayında yayınlamış bu araştırma 2004 sonrası yıllarda da bağımsız olarak uygulanmaya başlanmıştır. 2003 yılındaki araştırma ülkemizde mutluluk üzerine yapılmış ilk kapsamlı ve etkin araştırmadır. Bu araştırmanın temel amacı, Türkiye'deki bireylerin genel mutluluk düzeyine ilişkin algısını, yaşamdan ve kamu hizmetlerinden duyduğu memnuniyeti ölçmek ve bunların dönemler boyunca takibini yapmaktır. Yaşam Memnuniyeti araştırması 2003 yılından itibaren TÜİK tarafından Türkiye genelinde her yıl, 2013 yılından itibaren de il düzeyinde 3 yılda bir yapılan araştırmadır. Araştırmada mutluluk düzeyi, mutluluk kaynakları, umut düzeyi ve kamu hizmetlerinden memnuniyet düzeyi ölçülmektedir.

Aşağıdaki tabloda TÜİK'in 2003-2016 yılları arasında yapmış olduğu Yaşam Memnuniyeti Araştırması sonucunda, Türkiye genelinde insanların "Yaşamınızı bir bütün olarak düşündüğünüzde ne kadar mutlusunuz?" sorusuna verdikleri cevaplar gösterilmektedir.

Tablo 3. 2003-2016 Yılları Arası Mutluluk Düzeyi Değerleri

Mutluluk Düzeyi	YILLAR													
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Çok mutlu	12,0	9,3	9,1	8,8	8,7	8,2	7,7	9,2	8,5	8,5	9,6	8,1	7,9	7,6
Mutlu	47,6	48,7	48,5	49,1	51,5	47,5	46,6	52,0	53,6	52,5	49,5	48,2	48,6	53,8
Orta	33,2	29,8	29,5	30,3	28,8	30,3	31,1	28,1	28,0	28,9	30,2	32,0	32,0	28,3
Mutsuz	5,6	9,5	9,9	9,1	8,8	11,4	11,5	8,8	8,0	8,3	8,2	8,9	8,8	8,8
Çok mutsuz	1,7	2,6	2,9	2,8	2,2	2,5	3,1	2,0	1,9	1,8	2,6	2,8	2,6	1,6

Kaynak : (Türkiye İstatistik Kurumu, 2017)

Tablo incelendiğinde, yıllar geçtikçe kendini çok mutlu hisseden insanlarda azalma, mutlu hissedenlerde artış olduğu görülmektedir. Kendini mutlu hissedenlerde artış olduğu gibi yine yıllar geçtikçe kendini mutsuz hissedenlerde de artış gözlenmiştir. Kendini ne mutlu ne mutsuz (orta) hissedenler ve çok mutsuz hissedenlerde değişimin neredeyse olmadığı görülmektedir.

Türkiyede mutluluk ve yaşam memnuniyeti üzerine yapılmış çalışmalar incelendiğinde, çalışmaların Türkiye geneli değil, genellikle belirli bir grup (mesleki ya da sosyodemografik özelliklere göre) ile yapıldığı görülmektedir. Türkiye geneli yapılmış çalışmaların sayısı oldukça azdır. Bu çalışmalar ise aşağıdaki gibi özetlenebilir;

Üçdoğruk, Bozkuş ve Çevik, 2006 yılında yapmış oldukları “Bireysel Refah ve Mutluluk Düzeyine Etki Eden Faktörlerin Sıralı Lojit ile Modellenmesi : Türkiye Örneği” isimli çalışmalarında TÜİK Yaşam Memnuniyeti Araştırması 2004 yılı verilerini kullanarak belirleyici faktörleri Sıralı Lojit modelleri ile saptamışlardır. Çalışma sonuçlarına göre özet olarak; kadınların erkeklere göre daha mutlu olduğu, 45-60 yaş arasındaki bireylerin en mutsuz dönemlerini geçirdikleri, gelirin, sağlığın ve kentte yaşamının mutluluk düzeyini arttırdığı, eğitim düzeyinin artmasının ise mutluluğu azaltıcı bir etken olduğu belirlenmiştir (Bozkuş, Çevik, & Üçdoğruk, 2006).

Kahyaoglu, 2008 yılında yapmış olduğu “Yaşam Memnuniyeti ve Yaşam Memnuniyetini Etkileyen Değişkenler İle Ekonometrik Uygulama: Türkiye Örneği” isimli çalışmada, TÜİK Yaşam Memnuniyeti Araştırması 2004 yılı hane halkı anket verileri kullanılarak Sıralı Probit Modelleri kullanılmış olup sonuçlar iktisadi ve istatistiki olarak yorumlanmıştır. Çalışmanın sonucuna göre; kadınların erkeklere göre daha mutlu oldukları ve yaş ile mutluluk arasındaki ilişkinin U şeklinde olduğu görülmüştür. Yaşlı bireylerin beklentilerin aksine mutlu oldukları görülmüştür. Ayrıca gelirin, sağlığın ve refah düzeyindeki artışların mutluluk düzeyini arttırdığı belirlenmiştir. Kentte yaşamının mutluluk üzerinde artırıcı bir etkisi varken, eğitim düzeyinin artması mutluluğu artırıcı bir etken olarak belirlenmiştir (Kahyaoglu, 2008).

Babadağ, Uyanık ve Yılmaz, 2009 yılında yapmış oldukları “Yaşam Memnuniyeti Anketinin Veri Vadenciliği Süreci ile İncelenmesi” isimli çalışmalarında; TÜİK Yaşam Memnuniyeti Anketi 2003-2007 periyodu verilerini kullanmışlardır. Sonuç olarak,

evlilikten duyulan mutsuzluğun, umutsuzluğu ve mutsuzluğu tetiklediğini ve gelirin yüksek oluşunun ise umudu attırdığını ve mutluluğu olumlu yönde etkilediğini belirtmişlerdir (Köksal, 2011).

Giray, 2011 yılında yapmış olduğu “Doğrusal Olmayan Kanonik Korelasyon Analizi ve Yaşam Memnuniyeti Üzerine Bir Uygulama” isimli çalışmada TÜİK tarafından 2008 yılında yapılmış Yaşam Memnuniyet Anketi sonucunda ulaşılmış olan 6382 anket verisi kullanmıştır. Çalışmada Doğrusal Olmayan Kanonik Korelasyon yöntemi kullanılmış ve sonucunda; genç, evli ve ev kadınlarının; tüm yaşamlarını gözönünde bulundurduklarında yaşam memnuniyetlerinin orta üst mutluluk seviyesinde olduğu görülmüştür. Kentte ikamet eden, ortaokul mezunu, çalışan ve 700-900 TL arası gelire sahip olan erkekler de tüm yaşamlarını gözönünde bulundurduklarında yaşam memnuniyetlerinin orta üst mutluluk seviyesinde olduğu ve kendilerini en çok mutlu eden kişilerin tüm aileleri olduğu; ilkokul mezunu, orta yaşlı, emeklilerin de mutluluk seviyesinin orta-üst mutluluk seviyesinde olduğu ve kendilerini en mutlu eden kavramın sağlık olduğu sonucuna ulaşılmıştır (Giray, 2011).

Beşel, 2015 yılında yapmış olduğu; “2013 Yılı Yaşam Memnuniyeti Araştırması Sonuçlarının İl Bazlı Ekonomik, Sosyal ve Siyasi Analizi” isimli çalışmada TÜİK’in 2013 yılında yapmış olduğu Yaşam Memnuniyeti Araştırması verilerini kullanmıştır. Araştırmaya göre en mutlu ve en mutsuz illere ait göstergeler ekonomik, sosyal ve siyasi (2009-2014 yerel seçimleri) açıdan analiz edilmiştir. Sonuçlara göre; mutlu illerde mutsuz illere kıyasla; işsizlik oranları, konut satış sayısı oranları, genel bütçe vergi gelirlerinde toplam tahsilat içindeki payı, genel bütçe vergi gelirleri tahsilat/tahakkuk oranı, boşanma hızı, kişi başına elektrik tüketimi, nüfus yoğunluğu daha düşük iken; bin kişi başına otomobil sayısı, ev sahibi olan hanehalkı oranı, ilkokullarda net okullaşma oranı ve yüz bin kişi başına hastane yatak sayısı daha fazladır. Ayrıca 2014 yerel seçimleri sonuçlarına göre mutlu illerin belediye yönetiminde siyasi parti tercihi değişmemişken mutsuz illerin ikisinde belediye yönetiminin el değiştirdiği görülmüştür (Beşel, 2015).

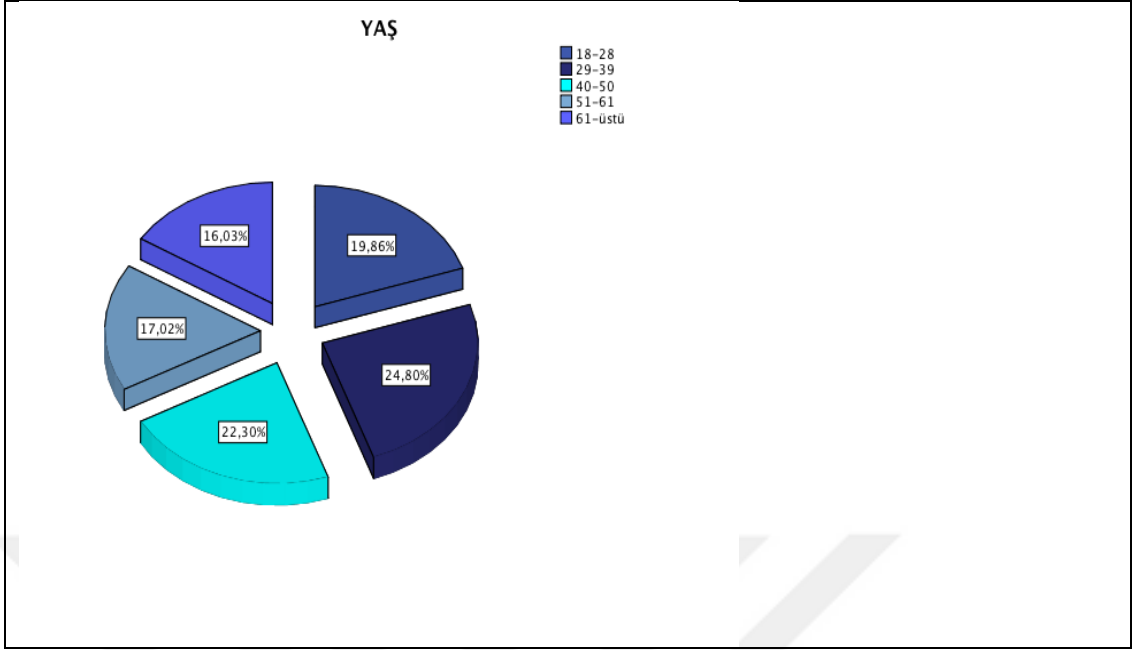
Yapılan literatür taramasında mutluluk üzerine yapılan araştırmaların mutluluk konusunu genel anlamda mutluluk derecesi olarak ele aldığı görülmüştür. Çalışmalar da mutluluk derecesi bağımlı değişken, sosyodemografik faktörler ise bağımsız değişken olarak kabul edilerek, mutluluk üzerinde etkili olan faktörlerin belirlenmesi amaçlandığı görülmüştür (Bülbül & Giray, 2011).

Bu çalışmada; TÜİK tarafından “Yaşam Memnuniyeti Anketi” kapsamında 2015 yılında tüm Türkiye’yi temsil edebilecek 9387 kişilik bir örneklem üzerinde toplanan veriler kullanılmıştır. Çalışmada yaşam memnuniyeti literatüründe kullanılan analiz tekniklerinin dışında Sınıflama ve Regresyon Ağacı teknikleri kullanılmıştır.

3.2 Uygulama

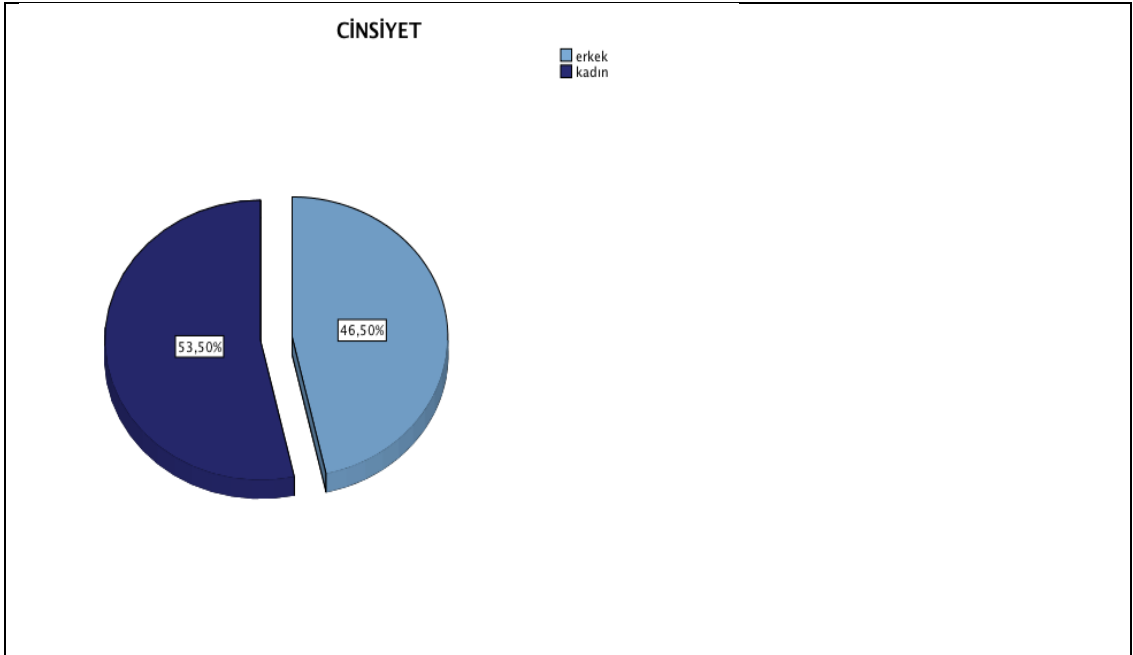
Uygulamanın birinci aşamasında analizde kullanılan değişkenlerin frekans dağılımları ve pasta grafikleri verilmiş ve kısaca yorumlanmıştır. Çalışmanın ikinci aşamasında “Yaşamınızı bir bütün olarak düşündüğünüzde ne kadar mutlusunuz?” sorusu bağımlı değişken olarak, 24 adet değişken ise bağımsız değişken (bağımsız değişkenlere konu ile ilgili uzman görüşleri alınarak karar verilmiştir) olarak alınarak sınıflama ağaçları oluşturulmuştur. Oluşturulan sınıflama ağaçlarında CART ve CHAID algoritmaları kullanılmıştır. Her iki algoritmada da veri setinin sırasıyla %70’i, %50’si ve %30’u başlangıç veri seti diğer bir adıyla eğitim verisi olarak ele alınmıştır. Buna göre veri setinin geri kalan %30, %50 ve %70’lik kısmı ise modeli test etmek için kullanılmıştır ve sonuçlar yorumlanmıştır. Uygulamanın üçüncü aşamasında, ankette sorulmuş olan memnuniyet sorularından toplam bir memnuniyet puanı oluşturulmuştur. Verilen cevaplar; çok memnun 5, memnun 4, orta 3, memnun değil 2, hiç memnun değil 1 olarak kodlanmıştır. Eğitiminden memnun olma değişkeninde “hiç eğitim almamışlar” 0, sağlık hizmetleri, asayiş hizmetleri, adli hizmetler, eğitim hizmetleri, Sosyal Güvenlik Kurumu’nun hizmetleri ve ulaşım hizmetleri değişkenlerinde “fikri yok” seçeneği de 0 olarak kodlanmıştır. Memnuniyet puanlarının yüksek olması yüksek memnuniyeti ifade etmektedir. Sayısal değer düştükçe memnuniyet değeride düşmektedir. Puanlar en düşük 13, en yüksek 105’tir. Bu memnuniyet puanı bağımlı değişken alınarak regresyon ağaçları oluşturulmuştur. Bağımsız değişkenler ise, cinsiyet, medeni durum, çalışılan kurum, mutluluk düzeyi, kişiyi kimin mutlu ettiği ve kişiyi neyin mutlu ettiği olarak belirlenmiştir. Regresyon ağaçlarında da sınıflama ağaçlarında olduğu gibi CART ve CHAID algoritmaları kullanılmıştır. Her iki algoritmada da veri setinin sırasıyla %70’i, %50’si ve %30’u başlangıç veri seti diğer bir adıyla eğitim verisi olarak ele alınmıştır. Buna göre veri setinin geri kalan %30, %50 ve %70’lik kısmı ise modeli test etmek için kullanılmıştır ve sonuçlar yorumlanmıştır.

3.2.1 Değişkenlere ait Tanımlayıcı İstatistikler



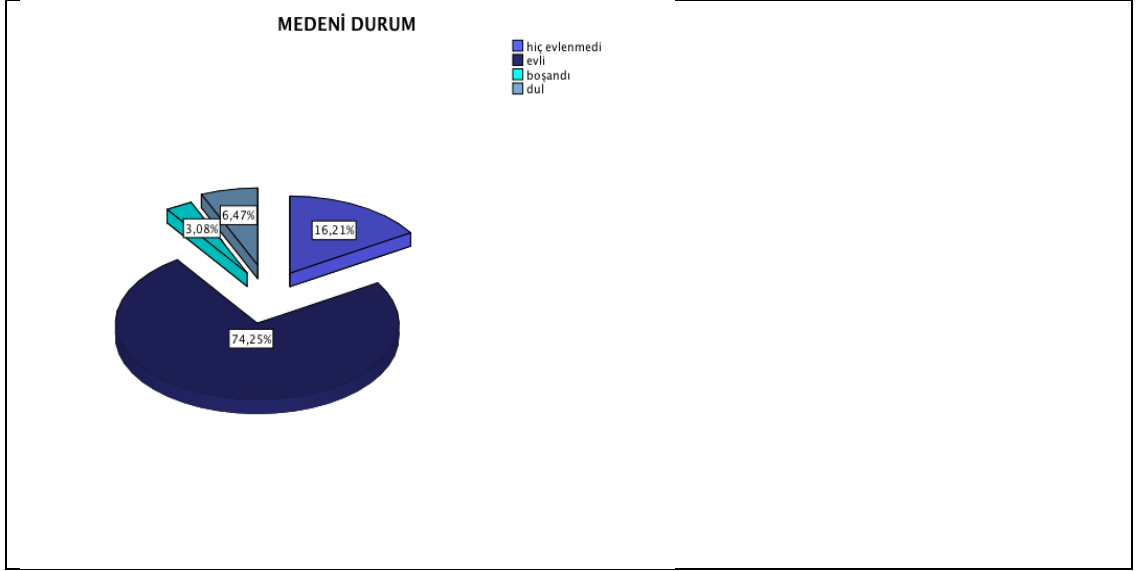
Şekil 3. Katılımcıların Yaş Dağılımı

TÜİK'in 9387 kişilik örneklem ile yaptığı çalışmada katılımcıların yaşları 18 ile 96 yaş arasındadır. Yaşlar 18-28, 29-39, 40-50, 51-61, 62 – 62üstü olarak gruplandırılmıştır. Sonuca göre yaş grupları dağılımları yaklaşık olarak benzerlik göstermekte ve katılımcıların yaklaşık %20'si 18-28 yaşları arasında, yaklaşık %25'i 29-39 yaşları arasında, yaklaşık %22'si 40-50 yaşları arasında, yaklaşık %17'si 51-61 yaşları arasında ve yaklaşık %16'sında 62 yaş ve üzeridir.



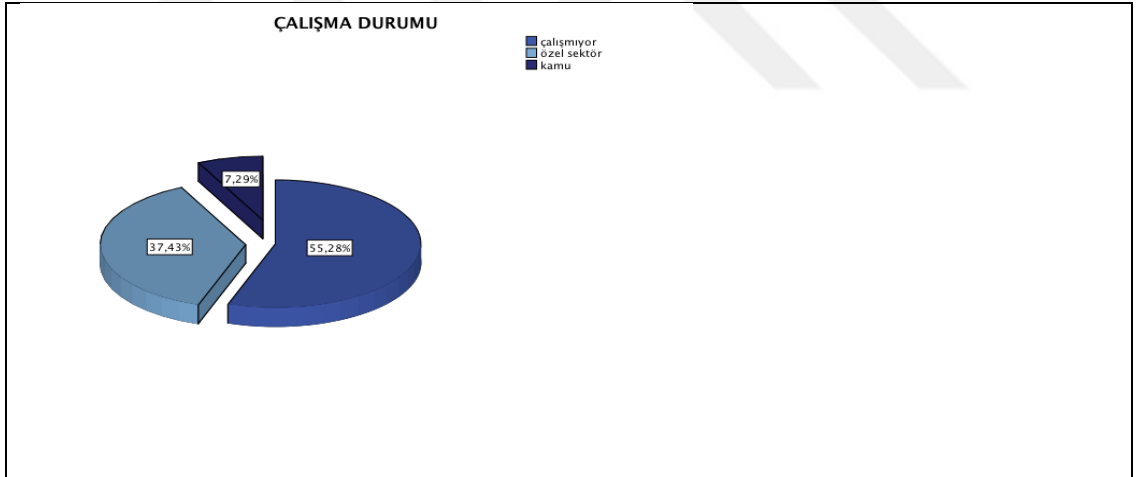
Şekil 4. Katılımcıların Cinsiyet Dağılımı

Katılımcıların cinsiyet dağılımı %46,5 erkek, %53,5 kadın şeklindedir.



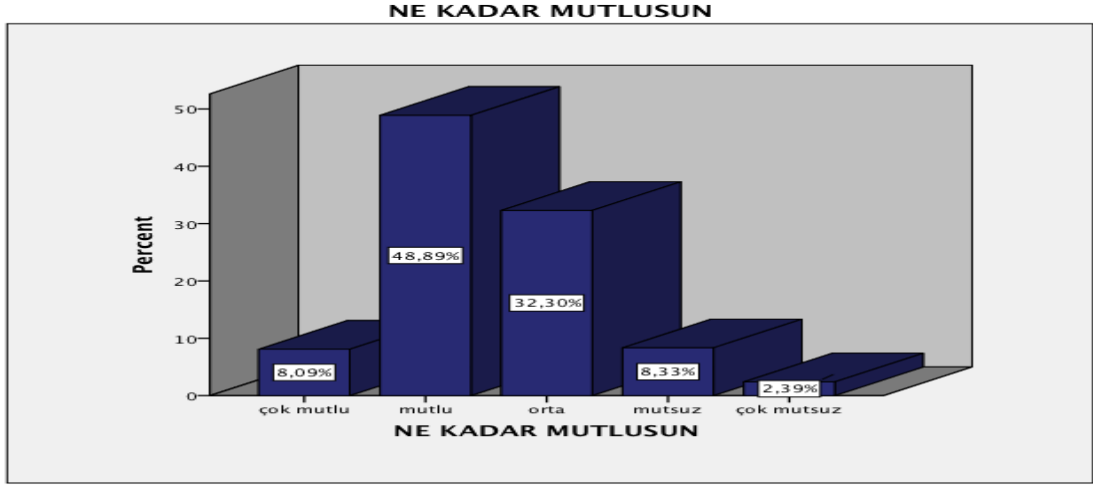
Şekil 5. Katılımcıların Medeni Durum Dağılımı

Katılımcıların medeni durumlarının dağılımlarının arasında büyük farkın olduğu şekil 5'tede görülmektedir. Dağılıma göre; katılımcıların yaklaşık %74'ü evli, yaklaşık %16'sı hiç evlenmemiş, yaklaşık %7'si dul, %3'ü ise boşanmış kişilerden oluşmaktadır.



Şekil 6. Katılımcıların Çalışma Durum Dağılımı

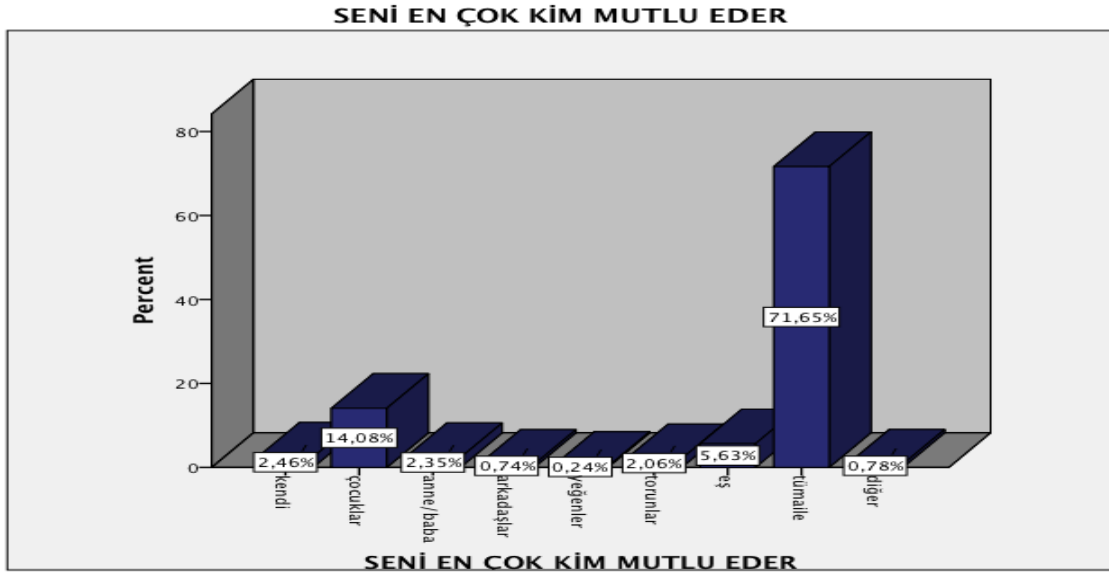
Katılımcıların çalışma durumlarının dağılımı şekil 6'da görülmektedir. Dağılıma göre; katılımcıların yaklaşık %55,3'ü çalışmıyor, yaklaşık %7,3'ü kamuda çalışıyor, yaklaşık %37,4'ü ise özel sektörde çalışmaktadır.



Şekil 7. Katılımcıların Mutluluk Düzeyi Dağılımı

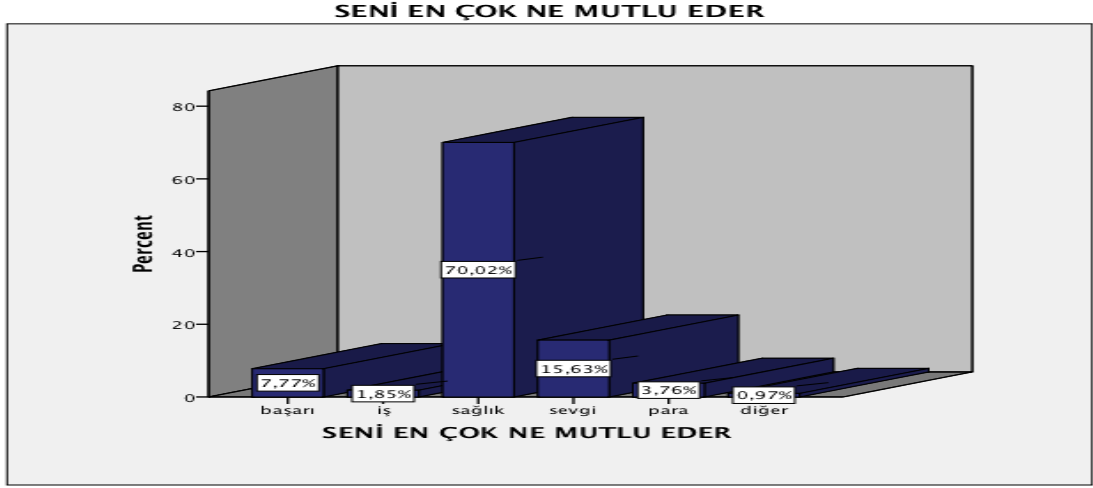
Katılımcıların yarısına yakın yaklaşık %49'unun kendimi mutlu, yaklaşık %32'sinin orta, %8'inin çok mutlu, yine yaklaşık olarak %8'inin mutsuz, yaklaşık %2'sinde çok mutsuz hissettiği şekil 7'de görülmektedir.

Bu sonuca göre 31'li değişken olan mutluluk düzeylerine bakıldığında katılımcıların yüksek oranda kendilerini mutlu ve orta hissettikleri, çok düşük oranlarda çok mutlu, mutsuz ve çok mutsuz hissettikleri görülmektedir.



Şekil 8. Katılımcıları En Çok Kimin Mutlu Ettiğinin Dağılımı

Katılımcıları en çok kimin mutlu ettiğine bakıldığında, katılımcıların yaklaşık %72'sinin tüm aile olduğu, %14'ünün çocuklar olduğu, yaklaşık %6'sının eş, yaklaşık %2'lik oranlarda da kendi, anne/baba, torunların olduğu, arkadaş, yeğenler ve diğer olarak belirtenlerin ise %1'lik bir oranı bile oluşturmadıkları şekil 8'de görülmektedir.



Şekil 9. Katılımcıları En Çok Neyin Mutlu Ettiğinin Dağılımı

Katılımcıları en çok kimin neyin mutlu ettiğine bakıldığında, katılımcıların yaklaşık %70'inin sağlık olduğu, %16'sının sevgi olduğu, yaklaşık %8'inin başarı, yaklaşık %4'ünün para olduğu, yaklaşık %2'sinin iş, %1'inin ise diğer olarak belirttiği başla bir nedenin olduğu şekil 9'da görülmektedir.

Tablo 4. Katılımcıların Mutlulukla İlişkilendirilen Konulardaki Memnuniyet Düzeyleri

	YÜZDE(%)					Toplam Yüzde
	Çok Memnun	Memnun	Orta	Memnun Değil	Hiç Memnun Değil	
Sağlığınızdan memnunmusun	6,8	60,8	18,1	12,2	2,0	100,0
Evliliğinizden memnunmusun	17,5	75,9	4,6	1,2	0,7	100,0
Oturduğunuz semtten memnunmusun	6,0	75,4	8,1	8,9	1,6	100,0
İşinizden memnunmusun	6,4	72,6	10,9	9,1	1,1	100,0
Gelirinizden memnunmusun	1,7	39,1	23,1	29,8	6,3	100,0
Sosyal hayatınızdan memnunmusun	1,7	43,3	22,8	27,1	5,1	100,0
Kendinize ayırdığınız zamandan memnunmusun	2,4	52,5	19,8	21,8	3,4	100,0

Arkadaşlarınız ile ilişkinizden memnunmusunuz	7,9	81,3	7,4	2,9	0,4	100,0
---	-----	------	-----	-----	-----	-------

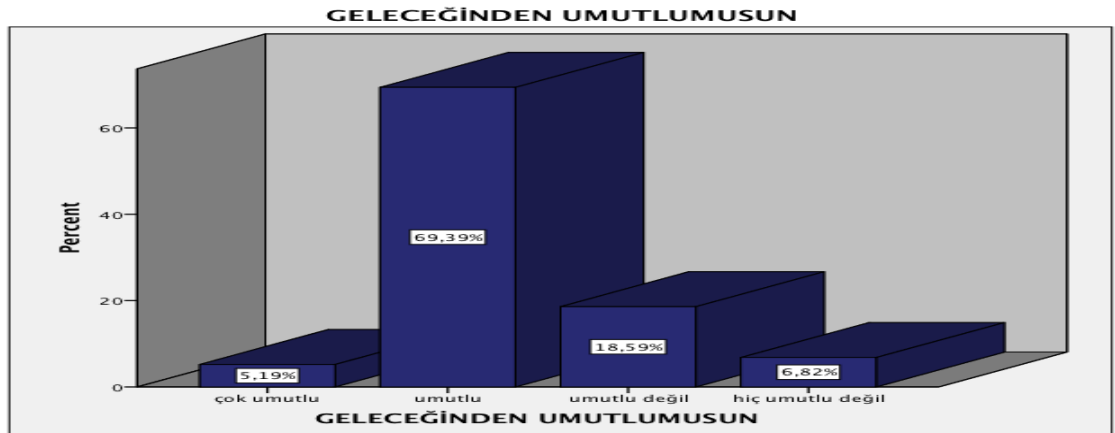
Tablo 5. Katılımcıların Bugüne Kadar Aldıkları Eğitimden Memnuniyet Düzeyleri

Aldığınız eğitimden memnunmusunuz	YÜZDE(%)						Toplam
	Çok memnun	Memnun	Orta	Memnun değil	Hiç memnun değil	Eğitim almadı	
	2,8	49,7	13,7	21,4	3,7	8,8	100,0

Tablo 4 ve Tablo 5'te katılımcıların sağlıklarından, evliliklerinden, aldıkları eğitimden, oturdukları semtlerden, işlerinden, gelirlerinden, sosyal hayatlarından, kendilerine ayırdıkları zamandan ve arkadaşları ile olan ilişkilerinden memnuniyet düzeylerinin sonuçları görülmektedir.

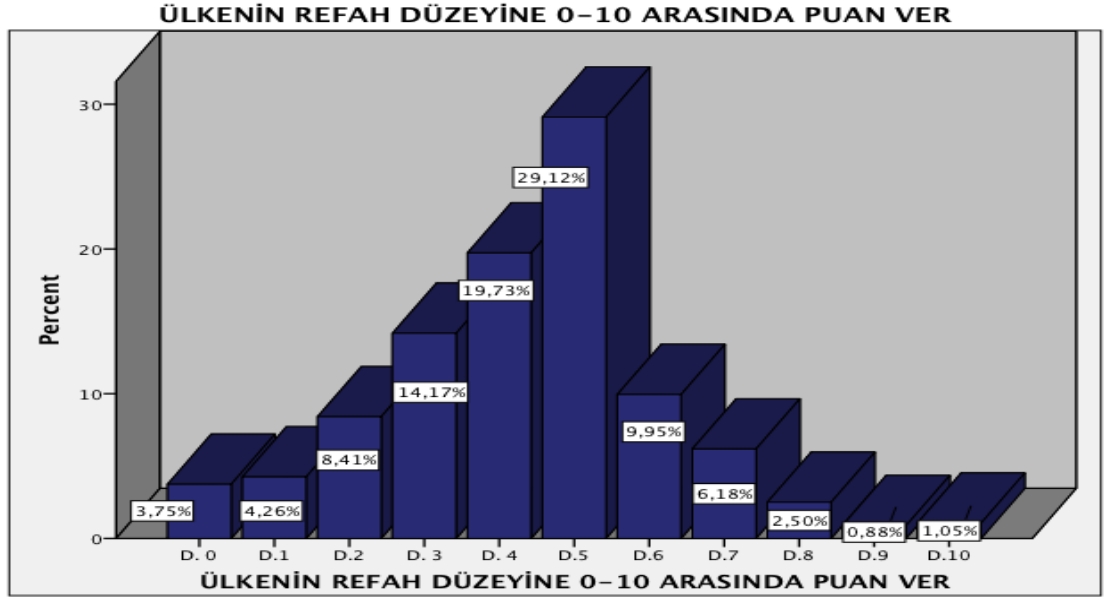
Tablo 4'teki sonuçlara göre katılımcıların evliliklerinden, oturdukları semtten, işlerinden ve arkadaşları ile olan ilişkilerinden memnun olanların oranı %70'in üzerinde, memnun olmayanların oranı %10'un altında olduğu görülmektedir. Katılımcıların gelirinden, sosyal hayatından ve kendine ayırdığı zamandan ise memnun olanların oranı yaklaşık %50'nin altında, memnun olmayanların oranı ise yaklaşık %30'a yakın seviyelerdedir. Katılımcıların %60'ı ise sağlığından memnun, %18'i orta seviyede, %12'si ise memnun değildir.

Tablo 5'in farklı olması eğitimden alınan memnuniyette hiç eğitim almamış kategorisinin de olmasından kaynaklanmaktadır. Tablo 5'e bakıldığı zaman katılımcıların yaklaşık olarak yarısı bugüne kadar aldıkları eğitimden memnundur. Yine Tablo 5'te katılımcıların küçümsenmeyecek kadar yaklaşık %9'luk kısmının bugüne kadar hiç eğitim almadığı görülmektedir.



Şekil 10. Katılımcıların Geleceklerinden Umutlu Olma Dağılımı

Katılımcıların umut düzeylerine bakıldığı zaman yaklaşık %69'unun gelecekte umutlu,yaklaşık %19'unun ise umutlu olmadığı Şekil 10'da görülmektedir.



Şekil 11. Katılımcıların Ülkenin Refah Düzeyine Göre Verdikleri Puan Dağılımı

Katılımcıların ülkenin refah düzeyine göre 0 ile 10 arası verdikleri puanların dağılımına göre oluşan grafik normal dağılım grafiği şeklinde oluşmuştur. Bu sonuca göre katılımcıların çoğunluğunu oluşturan yaklaşık %29'luk kısım ülkenin refah düzeyine 5 puan vermişlerdir. Oluşan şekle göre refah düzeyi puanları 5'ten 0'a doğru azaldığında da, 5'ten 10'a doğru çoğaldığında da katılımcıların oranlarının düştüğü görülmektedir.

Tablo 6.Katılımcıların Çeşitli Konulara Olan İlgi Düzeylerinin Dağılımı

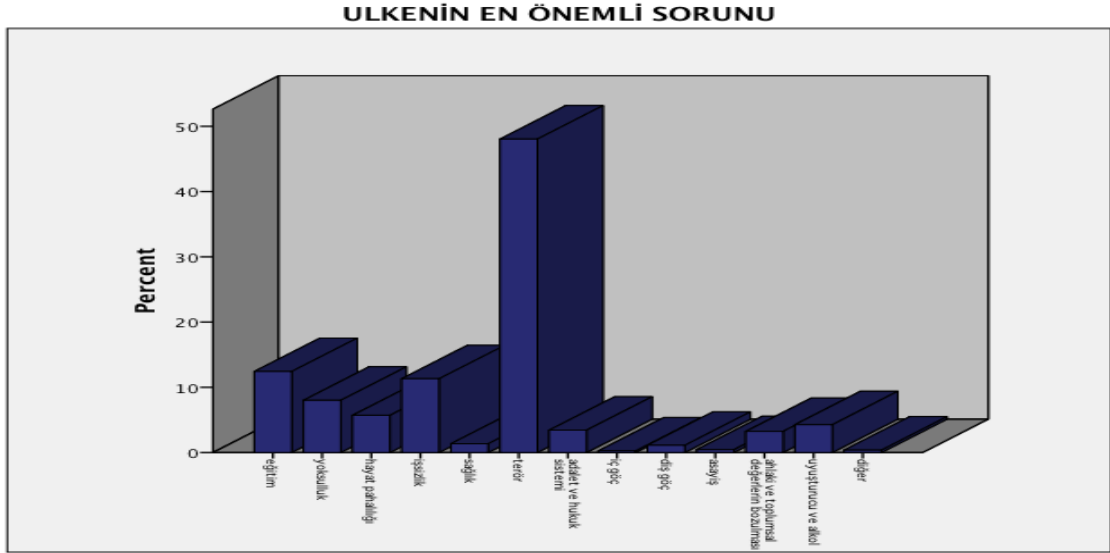
	YÜZDE(%)				Toplam
	İlgiliyim	Orta	İlgisizim	Fikrim Yok	
Bilim ile ilginiz	24,7	20,3	51,0	3,9	100,0
Kültür ile ilginiz	20,6	20,0	55,6	3,8	100,0
Ekonomi ile ilginiz	23,6	24,3	48,8	3,3	100,0
Siyaset ile ilginiz	20,4	24,8	52,1	2,7	100,0
Spor ile ilginiz	24,0	17,8	55,6	2,6	100,0
Moda ile ilginiz	9,8	14,0	73,5	2,7	100,0
Din ile ilginiz	66,4	22,0	10,9	0,7	100,0
Müzik ile ilginiz	37,3	25,6	35,4	1,7	100,0

Tablo 6’da katılımcıların bilim, kültür, ekonomi, siyaset, spor, moda, din, müzik konularında ilgi düzeylerinin sonuçları görülmektedir. Bu sonuçlara göre bir tek din konusunda katılımcıların yaklaşık %64’ünün ilgili olduğu, bilim, kültür, siyaset, spor konularında ilgili olanların %20’li oranlarda olduğu görülmektedir.

Moda en çok olmak üzere; bilim, kültür, ekonomi, siyaset, spor konularında ilgisizim diyenlerin oranının %50’nin üstünde çıktığı görülmektedir.

Tablo 7. Katılımcıların "Ülkenin en önemli sorunu nedir?" Görüşü

ÜLKENİN EN ÖNEMLİ SORUNU	FREKANS	YÜZDE(%)
Eğitim	1167	12,4
Yoksulluk	753	8,0
Hayat Pahalılığı	537	5,7
İşsizlik	1063	11,3
Sağlık	127	1,4
Terör	4515	48,0
Adalet ve hukuk sistemi	323	3,4
İç göç	24	0,3
Dış göç	105	1,1
Asayiş	42	0,4
Ahlaki ve toplumsal değerlerin bozulması	304	3,2
Uyuşturucu ve alkol	402	4,3
Diğer	35	0,4
Toplam	9397	100,0



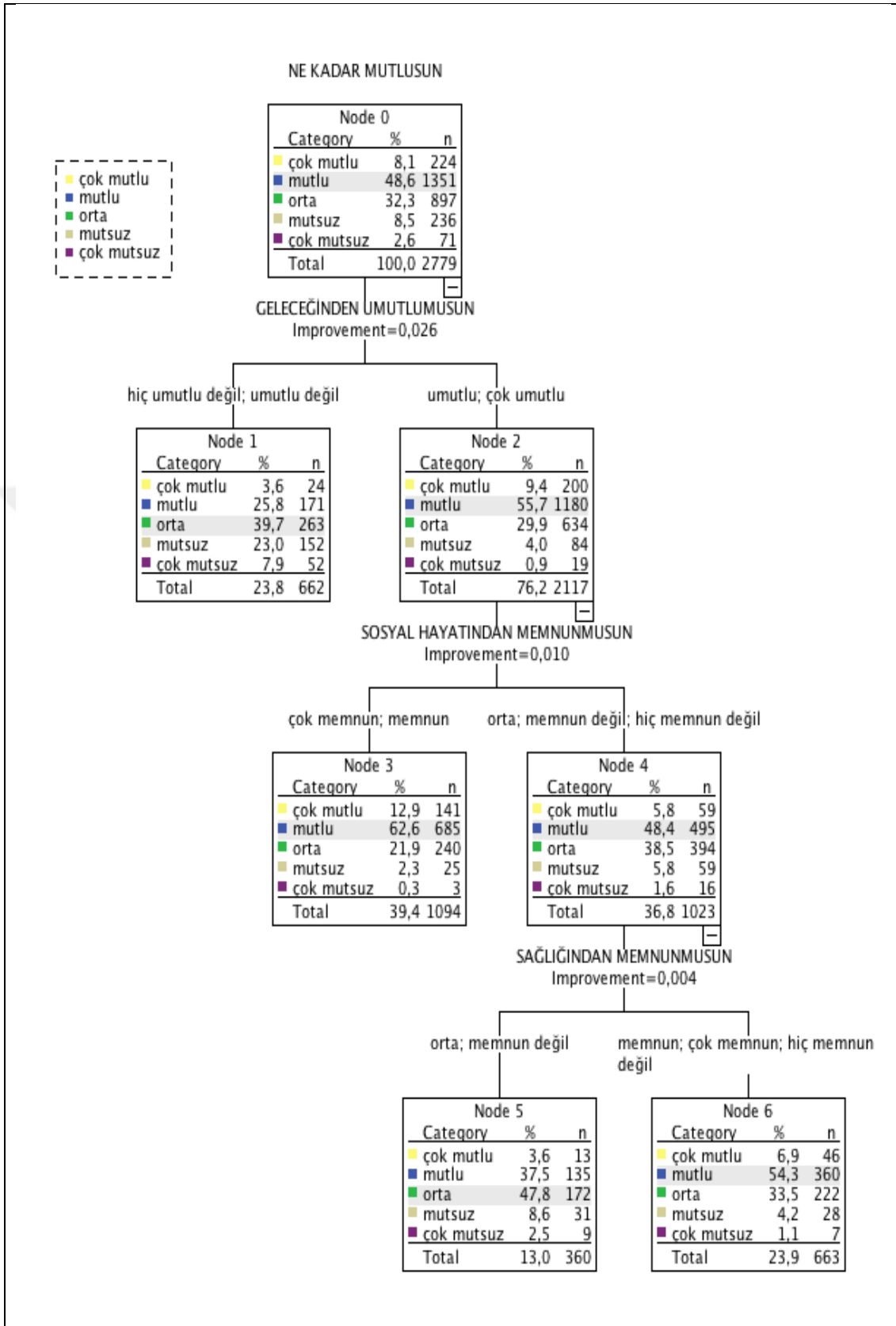
Şekil 12. Katılımcıların "Ülkenin en önemli sorunu nedir?" Grüşü

Katılımcılara; eğitim, yoksulluk, hayat pahalılığı, işsizlik, sağlık, terör, adalet ve hukuk sistemi, iç göç, dış göç, asayiş, ahlaki ve toplumsal değerlerin bozulması, uyuşturucu ve alkol, diğer seçenekleri sunulmuş ve bu seçenekler içerisinde "Sizce ülkenin önemli sorunu nedir?" diye sorulmuştur. Alınan cevaplar doğrultusunda katılımcıların yarısına yakın bir kısmı %48'i ülkenin en önemli sorununu terör olarak görmektedir. Terörden sonra en yüksek oran; katılımcıların yaklaşık %12 sine göre eğitim ve daha sonra katılımcıların yaklaşık %11'ine göre de işsizliktir.

3.2.2 CART Algoritması İle Oluşturulan Sınıflama Ağaçlarına Ait Bulgular

Başlangıçta CART algoritması ile modeli oluşturmak için veri setinin sırasıyla %70'i, %50'si ve %30'u başlangıç veri seti diğer bir adıyla eğitim verisi olarak ele alınmıştır. Buna göre veri setinin geri kalan %30, %50 ve %70'lik kısmı ise modeli test etmek için kullanılmıştır. Bu durumda 3 adet CART modeli elde edilmiştir.

CART algoritması ile kurulan modellerde Gini ayırma kriteri kullanılmıştır. İlk olarak veri setinin %30'u modeli oluşturmak, %70'i de modeli test etmek için kullanılmıştır.



Şekil 13. CART Algoritması ile Oluşturulan 1. Sınıflama Ağacı

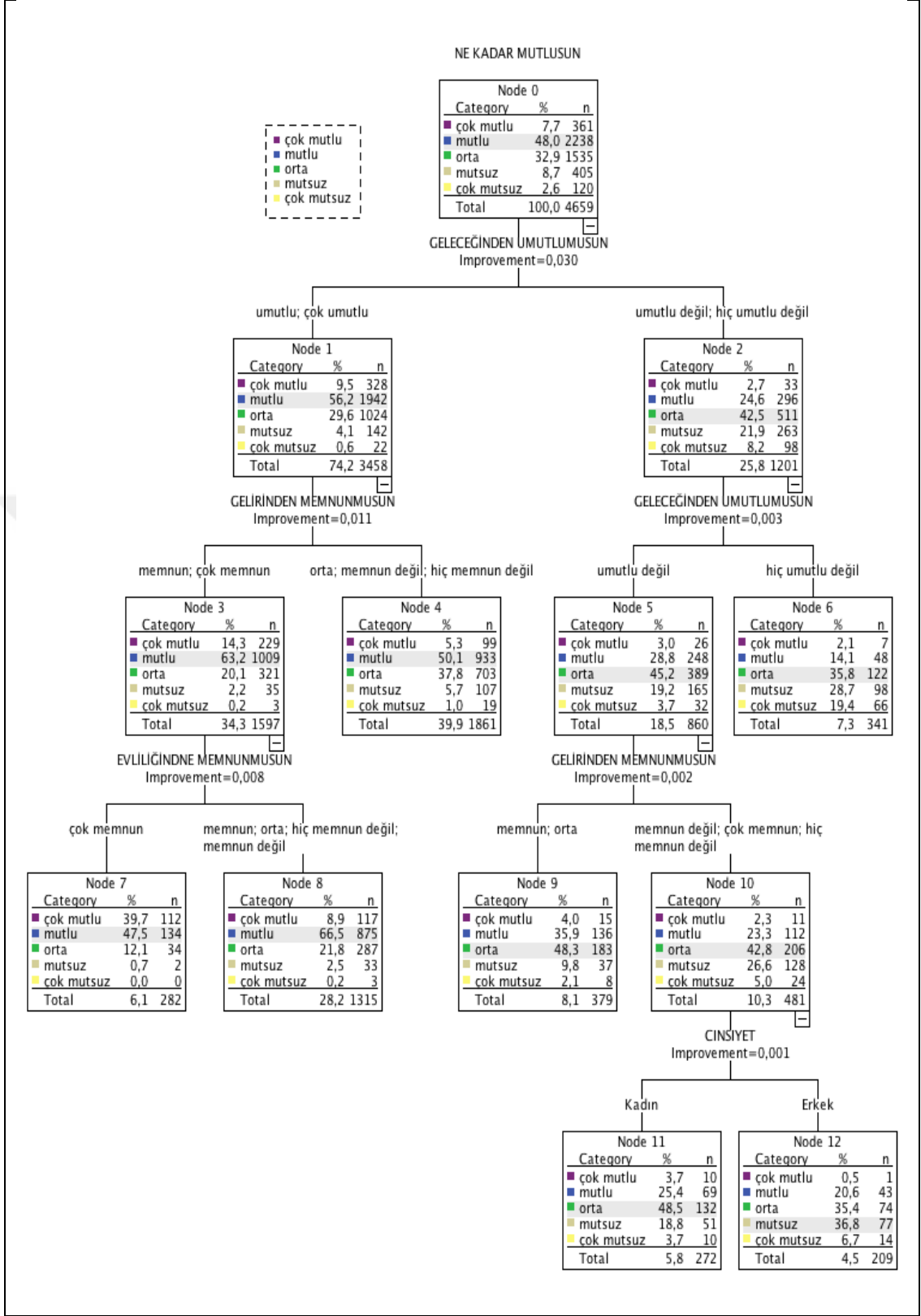
Şekil 13'te sunulan CART algoritması ile oluşturulan ilk sınıflama ağacı incelendiğinde, mutluluk düzeyleri üzerinde birinci derecede etkili bağımsız değişkenin gelecekte umutlu olma, ikinci derecede etkili bağımsız değişkenin sosyal hayatından memnun olma ve üçüncü derecede etkili bağımsız değişkenin sağlığından memnun olma olduğu belirlenmiştir. Sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya hiç umutlu değilim ve umutlu değilim 1 nolu sol çocuk düğümde, çok umutluyum ve umutluyum diyenler 2 nolu sağ çocuk düğümde sınıflanmıştır. Geleceğinden çok umutlu ve umutlu olanların %55,7'si mutlu, geleceğinden hiç umudu olmayanlar ile umudu olmayanların ise %25,7'i mutludur. Yine bu sonuca göre gelecekte umutlu olanların %4'ü mutsuzken, gelecekte umutlu olmayan kişilerin %23'ü mutsuzdur.

1 nolu çocuk düğüm saf halde yani terminal düğümdür. 2 nolu düğüm henüz saf değildir. 2 nolu düğümü saflaştırmak için ise sosyal hayatınızdan memnunsunuz sorusu kullanılmıştır. Bu soruya verilen yanıtlara 2. Düğüm iki alt düğüme (Düğüm3 ve Düğüm 4) ayrılmıştır. Düğüm 3'ü geleceğinden umutlu olan ve sosyal hayatından çok memnun olanlar ile memnun olanlar oluşturmaktadır. Düğüm 4'ü ise geleceğinden umutlu olan ve sosyal hayatından orta düzey memnun olan, memnun olmayan ile hiç memnun olmayanlar oluşturmaktadır. Düğüm 3'te geleceğinden umutlu olan ve sosyal hayatından çok memnun olanlar ile memnun olan insanların %62,6'sının mutlu olduğu sonucuna ulaşılmıştır. Düğüm 4'te geleceğinden umutlu olan ve sosyal hayatından orta derecede memnun olanlar, memnun olmayanlar ile hiç memnun olmayan insanların %48,4'ünün mutlu olduğu sonucuna ulaşılmıştır.

Düğüm 4'ü oluşturan geleceğinden umutlu olan ve sosyal hayatından orta derecede memnun olanlar, memnun olmayanlar ile hiç memnun olmayanlar sağlıklarından memnun olma durumlarına göre iki alt düğüme (Düğüm 5 ve Düğüm 6) ayrılmıştır. CART algoritması ile oluşturulan ilk sınıflama ağacı incelendiğinde mutluluğu etkileyen değişkenlerin gelecekte umutlu olmak, sosyal hayattan ve sağlıktan memnun olmak değişkenleri olduğu ve bu değişkenlerin arasında önemli bir ilişki olduğu sonucuna varılmıştır.

Mutluluğu etkileyen en önemli değişkenin gelecekte umutlu değişkeni olduğu sonucuna varılmıştır. Buna göre; umut seviyesi düştükçe mutluluk oranları da azalmakta olduğu sonucuna ulaşılmıştır. Aynı zamanda geleceğinden umutlu olup sosyal hayatından da memnun olan insanların sosyal hayatından memnun olmayan insanlara göre çok daha yüksek oranda mutlu olduğu sonucuna ulaşılmıştır.

CART algoritmasıyla oluşturulmuş ikinci ağaçta veri setinin yarısı modeli test etmek diğer yarısı modeli oluşturmak için kullanılmıştır. Veri setinin yarısı test etmek için, yarısı model oluşturmak için kullanıldığı zaman oluşan ağaç CART-2 ağacı olarak adlandırılmıştır.



CART algoritması ile oluşturulan ikinci sınıflama ağacı incelendiğinde, mutluluk düzeyleri üzerinde birinci derecede etkili bağımsız değişkenin gelecekte umutlu olma, ikinci derecede etkili bağımsız değişkenin gelirinden memnun olma, üçüncü derecede etkili bağımsız değişkenin evliliğinden memnun olma ve dördüncü dereceden etkili bağımsız değişkenin cinsiyet olduğu belirlenmiştir. Sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya umutluyum ve çok umutluyum diyen kişilerin 1 nolu sol çocuk düğümde, hiç umutlu değilim ve umutlu değilim diyen kişilerin 2 nolu sağ çocuk düğümde sınıflanmıştır. Geleceğinden çok umutlu ve umutlu olanların %56,2'ü mutlu, geleceğinden hiç umudu olmayanlar ile umudu olmayanların ise %24,6'i mutludur. Yine bu sonuca göre gelecekte umutlu olanların %4,1'u mutsuzken, gelecekte umutlu olmayan kişilerin yaklaşık %21,9'si mutsuzdur.

Şekil 14'te sunulan CART algoritması ile oluşturulmuş 2. sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya umutluyum ve çok umutlu diyenler 1 nolu sol çocuk düğümde, umutlu değilim ve hiç umutlu değilim diyenler 2 nolu sağ çocuk düğümde sınıflanmıştır. 1 ve 2 nolu düğümler henüz saf değildir.

1 nolu düğümü saflaştırmak için etki eden değişken katılımcılara sorulan gelirinizden memnunsunuz sorusudur. Bu soruya verilen yanıtlara göre; memnun olanlar ile çok memnun olanlar 3 nolu sol çocuk düğümde hiç memnun olmayanlar, memnun olmayanlar ve orta düzeyde memnun olanlar 4 nolu sağ çocuk düğümde sınıflanmıştır. 2 nolu düğümü saflaştırmak için ise yine geleceğizden umutlumsunuz sorusu kullanılmış ve verilen yanıtlara göre geleceğinden umutlu olmayanlar 5 nolu sol çocuk düğümünde, hiç umutlu olmayanlar ise 6 nolu sağ çocuk düğümde sınıflanmıştır. Oluşan sınıflamada 4 ve 6 nolu düğümler saflaşmış ve sınıflanmayı durdurmuş, 3 ve 5 nolu düğümler ise henüz saflaşmamıştır.

3 nolu düğümü saflaştırmak için etki eden değişken evliliğinizden memnunsunuz sorusudur. Evliliğinden çok memnun olanlar 7 nolu sol çocuk düğümde memnun olanlar, hiç memnun olmayanlar, memnun olmayanlar ile orta düzeyde memnun olmayanlar 8 nolu sağ çocuk düğümde sınıflanmışlardır. 7 ve 8 nolu düğümler saflaştığı için sınıflanma sonlanmıştır.

5 nolu düğümü saflaştırmak için etki eden değişken ise gelirinizden memnunsunuz sorusudur. Gelirinizden memnun olanlar ile orta derecede memnun olanlar 9 nolu sol çocuk düğümde gelirinizden memnun olmayanlar ise 10 nolu sağ çocuk düğümde sınıflanmışlardır. Sınıflama sonucu 9 nolu sol düğüm saflaştığı için uç düğüm haline gelmiş ve sınıflama durmuş; 10 nolu düğümde saflaşma meydana gelmediğinden saflaşma için cinsiyet değişkeni kullanılmıştır. Cinsiyet değişkeninde kadınlar 11 nolu sol çocuk düğümünde, erkekler ise 12 nolu sağ çocuk düğümünde sınıflandırılmıştır ve bu sınıflamalarda safsızlık meydana geldiği için sınıflama durmuştur.

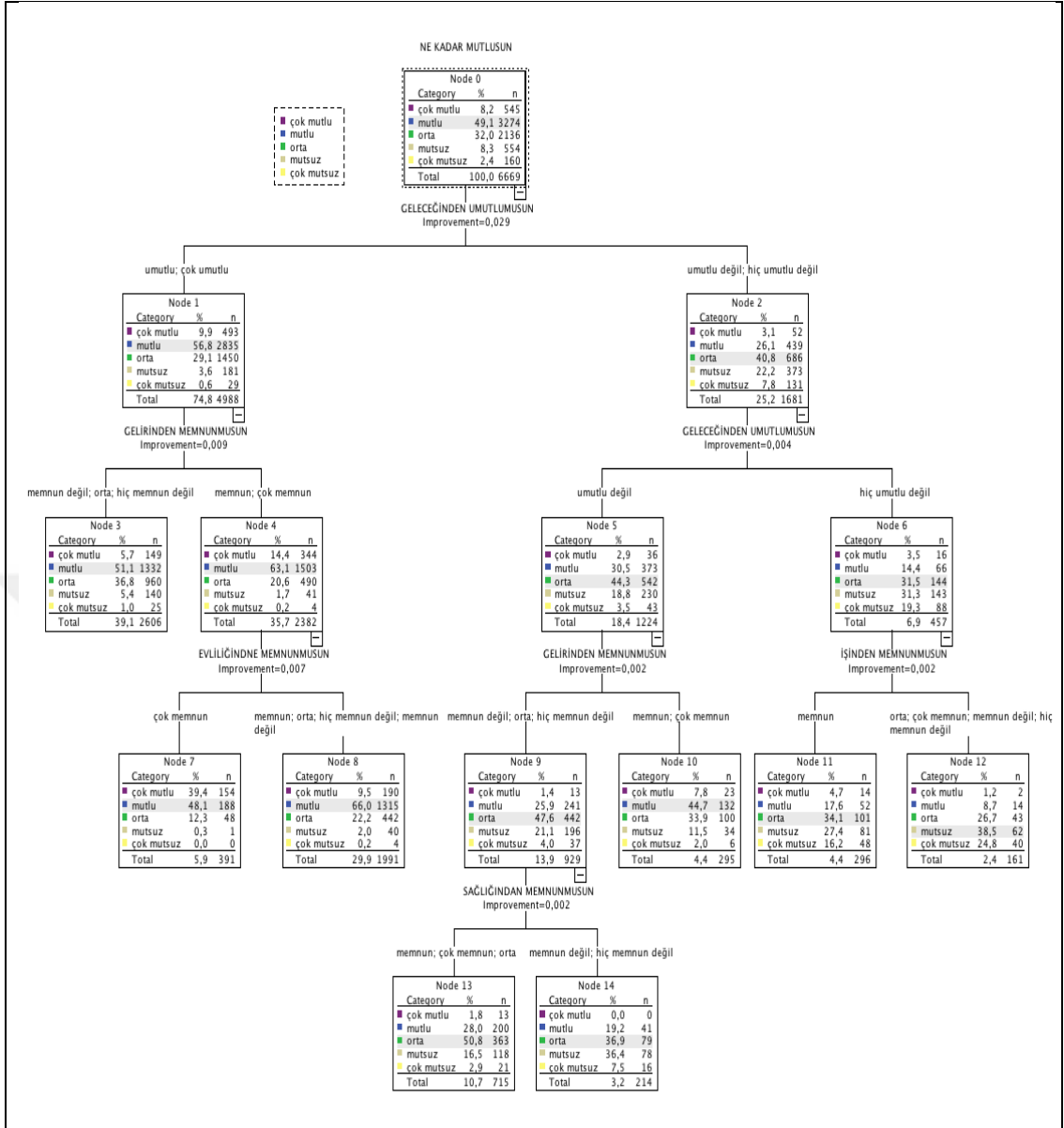
2. sınıflama ağacı incelendiğinde mutluluğu etkileyen değişkenlerin gelecekte umutlu olmak, gelirden, evlilikten ve cinsiyet değişkenleri olduğu ve bu değişkenlerin arasında önemli bir ilişki olduğu sonucuna varılmıştır.

Mutluluğu etkileyen en önemli değişkenin gelecekte umutlu olma değişkeni olduğu sonucuna varılmıştır. Bu sonuca göre gelecekte çok umutlu ve umutlu olanların yaklaşık %48'inin mutlu olduğu görülmektedir. Umut seviyesi düştükçe mutluluk oranları da önemli derecede azalmaktadır.

Gelecekte umutlu ve çok umutlu olanları gelirden memnun olma değişkeni etkilerken, gelecekte umutlu olmayanlar ile hiç umutlu olmayanlar önem derecesine göre kendi içinde tekrar sınıflanmıştır. Sınıflamaya göre gelecekte umutlu olmayan kişileri gelir değişkeninin etkilediği sonucuna varılmıştır.

CART algoritması ile oluşturulmuş ikinci sınıflama ağacına bakıldığı zaman gelecekte umutsuz olan kişileri hangi değişken etkilerse etkilesin hepsinin mutlu olmadığı görülmektedir. Mutluluk seviyeleri; geleceğinden umutlu olmayıp gelirden de memnun olmayan kadınlarda orta seviyede iken, hem geleceğinden umutlu olmayıp hemde gelirden memnun olmayan erkeklerde mutsuzluk seviyesi yüksektir.

CART algoritmasıyla oluşturulmuş üçüncü ağaçta veri setinin %70'i modeli oluşturmak, %30'u modeli test etmek için kullanılmıştır. Veri setinin yarısı test etmek için, yarısı model oluşturmak için kullanıldığı zaman oluşan ağaç CART-3 ağacı olarak adlandırılmıştır.



Node 12

Category	%	n
çok mutlu	1,2	2
mutlu	8,7	14
orta	26,7	43
mutsuz	38,5	62
çok mutsuz	24,8	40
Total	2,4	161

Şekil 15. CART Algoritması İle Oluşturulan 3. Sınıflama Ağacı

CART algoritması ile oluşturulan üçüncü sınıflama ağacı incelendiğinde, mutluluk düzeyleri üzerinde birinci derecede etkili bağımsız değişkenin gelecekte umutlu olma, ikinci derecede etkili bağımsız değişkenin gelirinden memnun olma, üçüncü derecede etkili bağımsız değişkenin evliliğinden memnun olma ve dördüncü dereceden etkili bağımsız değişkenin sağlığından memnun olma olduğu belirlenmiştir. Sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya umutluyum ve çok umutluyum diyen kişilerin 1 nolu sol çocuk düğümde, hiç umutlu değilim ve umutlu değilim diyen kişilerin 2 nolu sağ çocuk düğümde sınıflanmıştır. Geleceğinden çok umutlu ve umutlu olanların %56,8'ü mutlu, geleceğinden hiç umudu olmayanlar ile umudu olmayanların ise %26,1'i mutludur. Yine bu sonuca göre gelecekte umutlu olanların %3,6'u mutsuzken, gelecekte umutlu olmayan kişilerin yaklaşık %22,2'si mutsuzdur.

Şekil 15'te sunulan üçüncü sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya umutluyum ve çok umutlu diyenler 1 nolu sol çocuk düğümde, umutlu değilim ve hiç umutlu değilim diyenler 2 nolu sağ çocuk düğümde sınıflanmıştır. 1 ve 2 nolu düğümler henüz saf değildir.

1 nolu düğümü saflaştırmak için etki eden değişken katılımcılara sorulan gelirinizden memnunsunuz sorusudur. Bu soruya verilen yanıtlara göre; memnun olanlar ile çok memnun olanlar 3 nolu sol çocuk düğümde hiç memnun olmayanlar, memnun olmayanlar ve orta düzeyde memnun olanlar 4 nolu sağ çocuk düğümde sınıflanmıştır. Gelirinden memnun olan ve çok memnun olan gelecekte umutlu kişilerin(Düğüm 4) %63,1'i mutlu, gelirinden orta düzeyde memnun olan, memnun olmayan ve hiç memnun olmayan gelecekte umutlu kişilerin ise %51,1'inin mutlu olduğu sonucuna ulaşılmıştır. Bu sonuca göre gelirden memnun olan kişilerin daha mutlu olduğuna ulaşılabilmektedir.

2 nolu düğümü saflaştırmak için ise yine geleceğinizden umutlumsunuz sorusu kullanılmış ve verilen yanıtlara göre geleceğinizden umutlu olmayanlar 5 nolu sol çocuk düğümünde, hiç umutlu olmayanlar ise 6 nolu sağ çocuk düğümde sınıflanmıştır. Geleceğinizden umutlu olmayanların(Düğüm 5) %30,5'i, Geleceğinizden hiç umutlu olmayanların ise sadece %14,4'ünün mutlu olduğu sonucuna ulaşılmıştır. Oluşan sınıflamada 3 nolu düğüm saflaşmış ve sınıflanmayı durdurmuş, 4,5 ve 6 nolu düğümler ise henüz saflaşmamıştır. 4. düğümü saflaştırmak için etki eden değişken evliliğinizden memnunsunuz sorusudur. Evliliğinizden çok memnun olanlar 7 nolu sol çocuk düğümde memnun olanlar, hiç memnun olmayanlar, memnun olmayanlar ile orta düzey memnun olmayanlar 8 nolu sağ çocuk düğümde sınıflanmışlardır. 7 ve 8 nolu düğümler saflaştığı için sınıflanma sonlanmıştır. 5. düğümü saflaştırmak için etki eden değişken ise gelirinizden memnunsunuz sorusudur. Gelirinizden orta derecede memnun olanlar memnun olmayanlar, hiç memnun olmayanlar 9 nolu sol çocuk düğümde, gelirinizden memnun olanlar ile çok memnun olanlar ise 10 nolu sağ çocuk düğümde sınıflanmışlardır. Sınıflama sonucu 10 nolu sol düğüm saflaştığı için uç düğüm haline gelmiş ve sınıflama durmuş; 9 nolu düğümde saflaşma meydana gelmediğinden saflaşma için sağlığından memnun olma değişkeni kullanılmıştır. Sağlığından orta derecede memnun olan, memnun olan ve çok memnun olanlar 13 nolu sol çocuk düğümünde, memnun olmayanlar ve hiç memnun olmayanlar ise 14 nolu sağ çocuk düğümünde sınıflandırılmıştır ve bu sınıflamalarda safsızlık meydana geldiği için sınıflama durmuştur.

6. düğüm olan geleceğinizden hiç memnun olmayanlar işinden memnun olma değişkeni bakımından iki alt düğüme(Düğüm 11 ve Düğüm 12) ayrılmıştır. Düğüm 11 geleceğinizden hiç umutlu olmayan ama işinden memnun olan kişilerden oluşmaktadır. Bu kişilerin %17,6'sı mutludur. Düğüm 12'yi ise geleceğinizden hiç memnun olmayanlar ve işinden orta derecede memnun olanlar, memnun olmayanlar ve hiç memnun olmayanlar oluşturmaktadır. Bu kişilerin ise sadece %8,7'si mutludur.

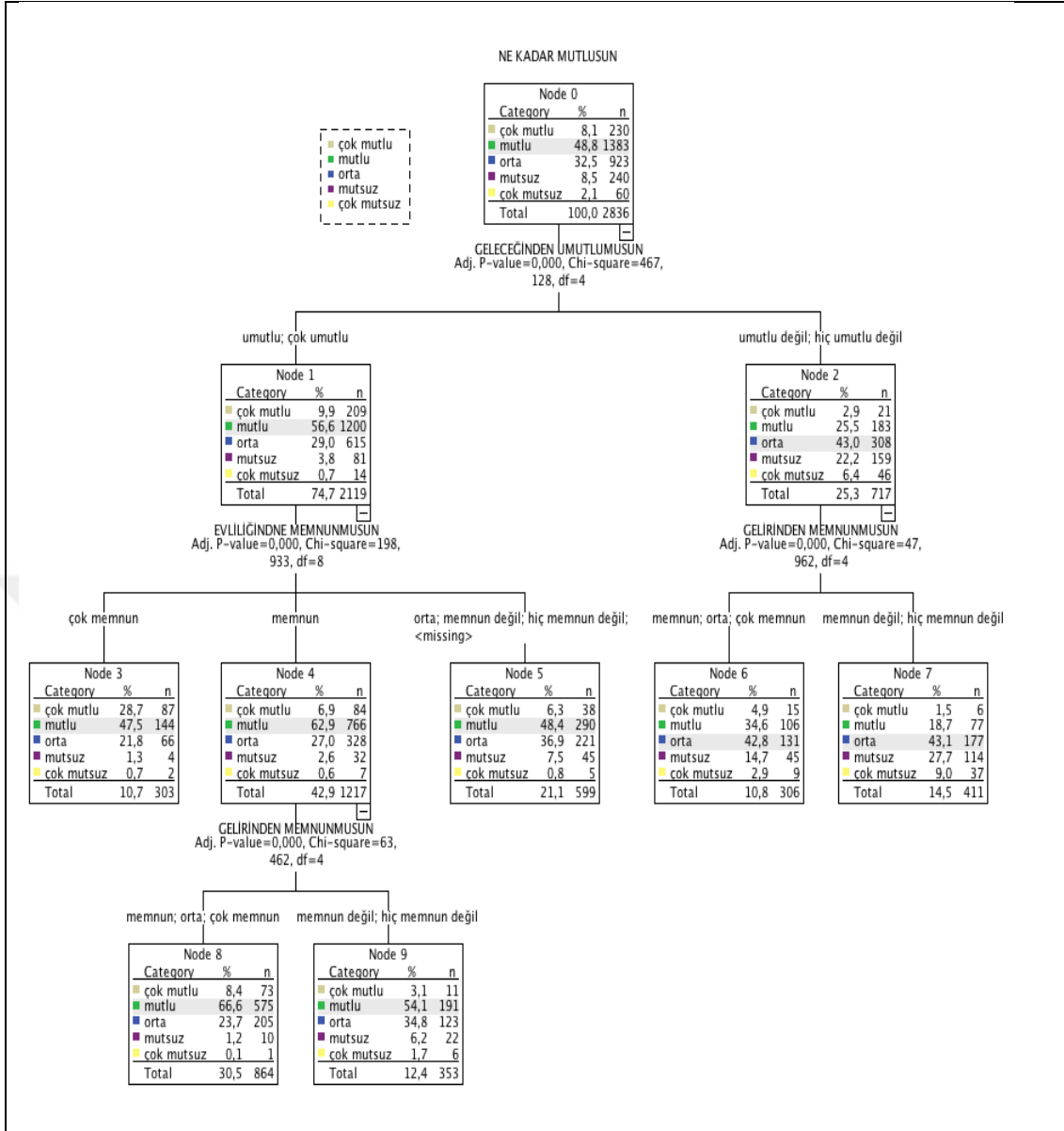
CART algoritması ile oluşturulmuş üçüncü sınıflama ağacına bakıldığı zaman gelecekte umutlu olmanın mutlu olma faktörü olduğu sonucuna ulaşılmıştır. Geleceğinizden umutlu olup hem gelirinizden hemde evliliğinizden memnun

olan mutlu insanların oranı %66 olduđu sonucuna ulařılırken, geleceđinden hi umutlu olmayan insanlar iřlerinden memnun olsalar bile mutlu insanların yzdesi 17,6'dır.

Sonu olarak bütun sınıflama ađaları incelendiđinde mutluluđu etkileyen en önemli deđiřkenin geleceklerinden umutlu olma deđiřkeni olduđu belirlenmiřtir. Fakat model verisi arttika mutluluđu etkileyen deđiřken sayısı ve dğüm sayısı da artmaktadır ve ađa daha karmařık bir yapıya bürünmektedir. Dolayısıyla, kategoriler yani sınıflar hakkında daha ayrıntılı bilgiye ulařılmaktadır. Sınıflar arasındaki etkileřimlerde görülebilmektedir. Bu durumda da arařtırmacı analiz ettiđi veriye ait daha detaylı bilgiye sahip olacađından model verisinin yüksek tutulması önerilmektedir.

3.2.3 CHAID Algoritması İle Oluřturulan Sınıflama Ađalarına Ait Bulgular

CHAID algoritması ile model oluřturulurken ilk olarak veri setinin %30'u modeli oluřturmak, %70'i de modeli test etmek için kullanılmıřtır. Bunun sonucunda oluřturulan ađa CHAID-1 olarak adlandırılmıř ve ařađıda gösterildiđi řekildedir.



Şekil 16. CHAID Algoritması İle Oluşturulan 1. Sınıflama Ağacı

CHAID algoritması ile oluşturulan ilk sınıflama ağacı incelendiğinde, mutluluk düzeyleri üzerinde birinci derecede etkili bağımsız değişkenin gelecekte umutlu olma, ikinci derecede etkili bağımsız değişkenin gelirinden memnun olma ve evliliğinden memnun olma, üçüncü derecede etkili bağımsız değişkenin gelirinden memnun olma olduğu belirlenmiştir. Sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya umutluyum ve çok umutluyum diyen kişilerin 1 nolu sol çocuk düğümde, hiç umutlu değilim ve umutlu değilim diyen kişilerin 2 nolu sağ çocuk düğümde sınıflanmıştır. Geleceğinden çok umutlu ve umutlu olanların %56,6'ü mutlu, geleceğinden hiç umudu olmayanlar ile umudu olmayanların ise %25,5'i mutludur. Yine bu sonuca göre gelecekte umutlu olanların %3,8'u mutsuzken, gelecekte umutlu olmayan kişilerin yaklaşık %22,2'si mutsuzdur.

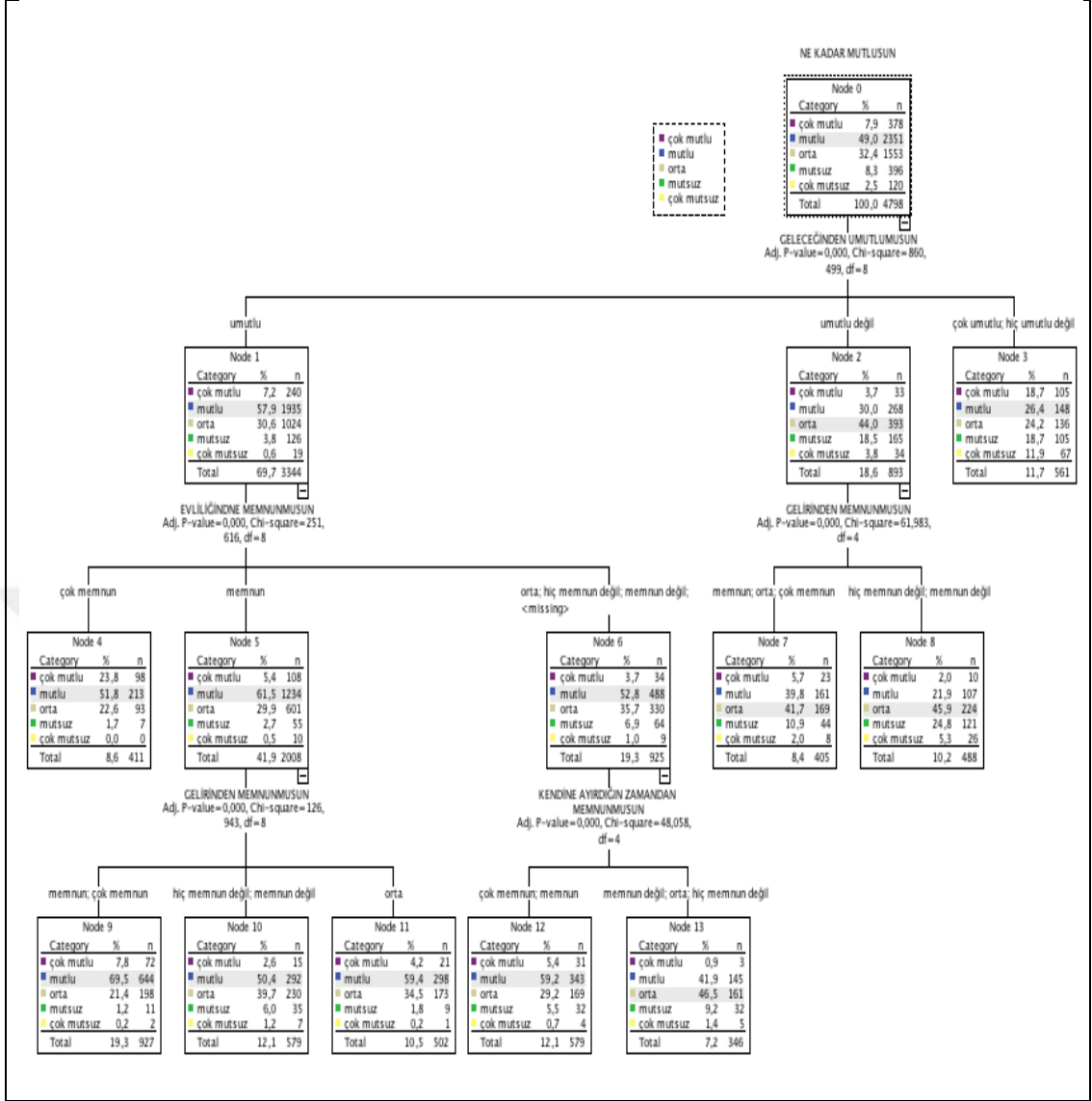
Mutluluk deęişkeni aile düęümünü ayıran ilk ayıraç gelecekte umutlu olunup olunmadığı sorusudur. Bu soruya umutluyum ve çok umutlu diyenler 1 nolu sol çocuk düęümde, umutlu deęilim ve hiç umutlu deęilim diyenler 2 nolu saę çocuk düęümde sınıflanmıştır.1 ve 2 nolu düęümler henüz saf deęildir. 1 nolu düęümü saflaştırmak için etki eden deęişken katılımcılara sorulan evlilięinizden memnunuz sorusudur.Bu soruya verilen yanıtlara göre; çok memnun olanlar 3 nolu sol çocuk düęümde, memnun olanlar 4 nolu ortanca çocuk düęümde, hiç memnun olmayanlar, memnun olmayanlar ve orta düzeyde memnun olanlar 5 nolu saę çocuk düęümde sınıflanmıştır. 3 ve 5 nolu düęümlerde saflaşma meydana geldiğinden uç düęüm haline gelmişler ve sınıflama durdurulmuşken 4 nolu düęüm saflaşmadığı için tekrardan sınıflandırılmıştır. 4 nolu düęümün saflaşması için gelirinizden memnunuz sorusu kullanılmış ve gelirinden çok memnun olanlar,memnun olanlar ve orta düzeyde memnun olanlar 8 nolu sol düęümde; memnun olmayanlar ile hiç memnun olmayanlar 9 nolu saę düęümde sınıflanmıştır. 8. Düęümü oluşturan geleceğinden umutlu, evlilięinden ve gelirinden memnun ve çok memnun olan insanların %66,6'sı, 9. Düęümü oluşturan geleceğinden umutlu, evlilięinden memnun fakat gelirinden memnun olmayan insanların ise %54,1'i mutludur. 8 ve 9 nolu düęümlerde saflaşma meydana geldiği için uç düęüm haline gelmiş sınıflama durmuştur.

2 nolu düęümü saflaştırmak için ise gelirinizden memnunuz sorusu kullanılmış ve verilen yanıtlara göre gelirinden çok memnun olanlar, memnun olanlar ve orta düzeyde memnun olanlar 6 nolu sol çocuk düęümünde, hiç memnun olmayanlar ve memnun olmayanlar ise 7 nolu saę çocuk düęümde sınıflanmıştır. 6 ve 7 nolu düęümlerde saflaşma gerçekte olduğu için uç düęüm haline gelmişler ve sınıflama durmuştur.

CHAID-1 ağacı incelendiğinde mutluluęu etkileyen deęişkenlerin gelecekte umutlu olmak, evlilikten ve gelirinden memnun olmak deęişkenleri olduğu ve bu deęişkenlerin arasında önemli bir ilişki olduğu sonucuna varılmıştır. Mutluluęu etkileyen en önemli deęişkenin gelecekte umutlu deęişkeni olduğu sonucuna varılmıştır.Umut seviyesi düştükçe mutluluk oranları da önemli derecede azalmaktadır.

Gelecekte umutlu ve çok umutlu olanları evlilięinden memnun olma deęişkeni etkilerken, gelecekte umutlu olmayanlar ile hiç umutlu olmayanları gelirinden memnun olma deęişkeni etkilemektedir. Geleceğinden umutlu olup evlilięinden memnun olan kişilerin yaklaşık %63'u mutlu kişiler, geleceğinden umutlu olmayıp gelirinden memnun olmayan kişilerin yaklaşık %43'ü ise mutluluk düzeyinde kendini orta düzey hisseden kişilerdir.

CHAID algoritması ile ikinci model oluşturulurken veri setinin yarısı modeli oluşturmak, yarısında modeli test etmek için kullanılmıştır.Bunun sonucunda oluşturulan ağaç şekil 17'de gösterilmiştir.



Şekil 17. CHAID Algoritması İle Oluşturulan 2. Sınıflama Ağacı

CHAID algoritması ile oluşturulan ikinci sınıflama ağacı incelendiğinde, mutluluk düzeyleri üzerinde birinci derecede etkili bağımsız değişkenin gelecekte umutlu olma, ikinci derecede etkili bağımsız değişkenin gelirinden memnun olma ve evliliğinden memnun olma, üçüncü derecede etkili bağımsız değişkenin gelirinden memnun olma ve kendine ayırdığı zamandan memnun olma olduğu belirlenmiştir. Sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya umutluyum diyen kişilerin 1 nolu sol çocuk düğümde, umutlu değilim diyen kişiler 2 nolu orta çocuk düğümde, çok umutluyum ve hiç umutlu değilim diyenler 3 nolu sağ çocuk düğümde sınıflanmıştır. Geleceğinden umutlu olanların %57,9'u mutlu, geleceğinden umutlu olmayanların ise %30'u mutludur. Geleceğinden çok umutlu olanlar ve hiç umutlu olmayanların ise %26,4'ü mutludur.

Geleceğinden umutlu olan kişilerin oluşturduğu 1. Düğüm evliliğinden memnun olma durumuna göre 3 alt düğüme(Düğüm 4, 5 ve 6) ayrılmıştır.Bu soruya verilen yanıtlara göre; çok memnun olanlar 4 nolu sol çocuk düğümde, memnun olanlar 5 nolu ortanca çocuk düğümünde, memnun olmayanlar, hiç memnun olmayanlar ve orta derecede memnun olanlar 6 nolu sağ çocuk düğümde sınıflanmıştır. Sınıflamada 4 nolu düğüm saflaştığı için uç düğüm haline gelmiş ve sınıflama durmuşken, 5 ve 6 nolu düğümler henüz saf değildirler.

Geleceğinden umutlu olup evliliğinden memnun olanların oluşturduğu 5. Düğüm gelirden memnun olma durumuna göre 3 alt düğüme (Düğüm 9, 10 ve 11) ayrılmıştır.Geleceğinden umutlu olup evliliğinden memnun ve gelirinden memnun ve çok memnun olan mutlu insanların oranı %69,5, geleceğinden umutlu olup evliliğinden memnun ve gelirinden orta derecede memnun olan mutlu insanların oranı %59,4, geleceğinden umutlu olup evliliğinden memnun ve gelirinden hiç memnun olmayan ve memnun olmayan mutlu insanların oranı %50,4 olduğu sonucuna ulaşılmıştır.

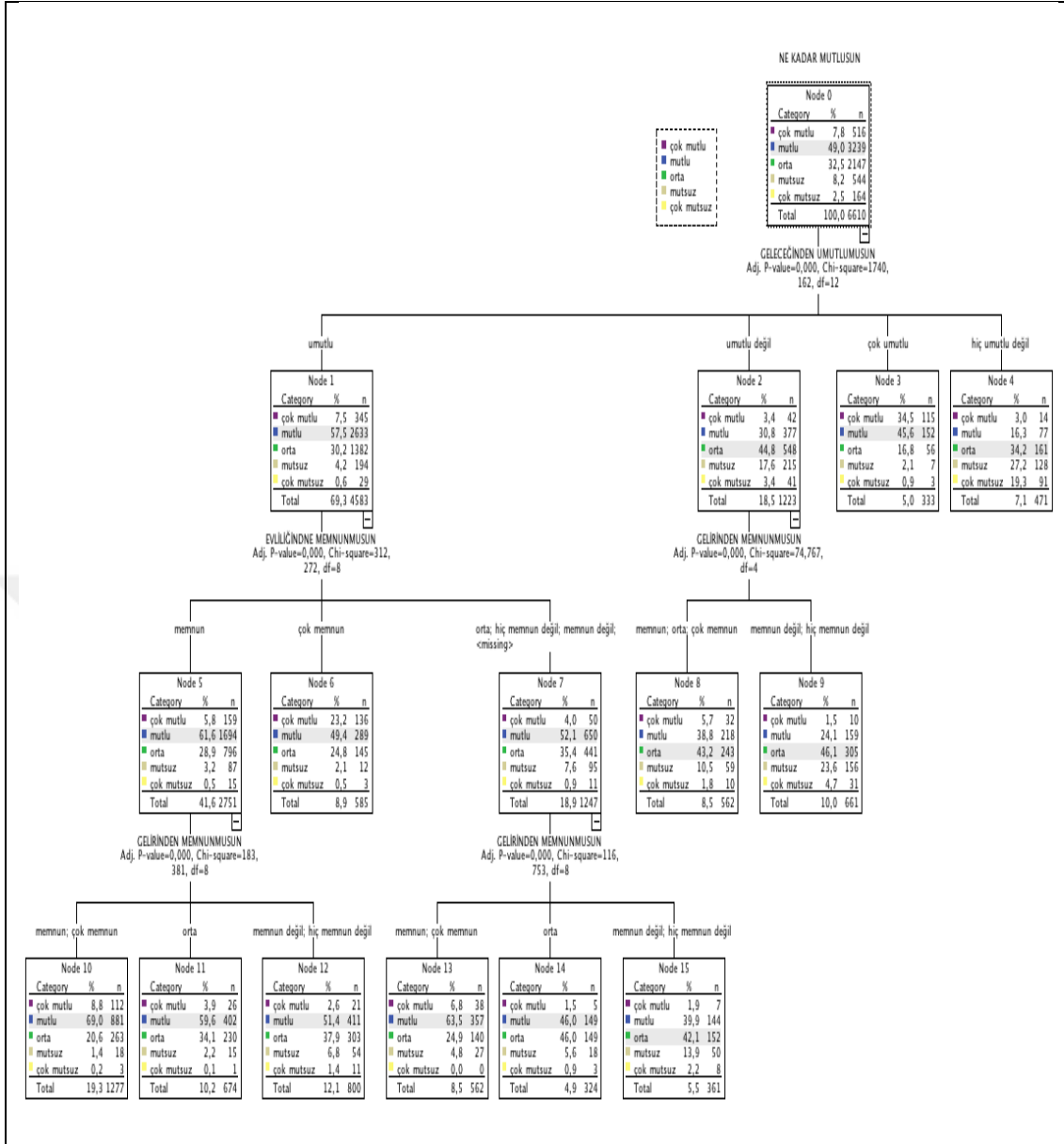
Geleceğinden umutlu olup evliliğinden orta derecede memnun olan, memnun olmayan ve hiç memnun olmayanların oluşturduğu 6. Düğüm kendine ayırdığı zamandan memnun olma durumuna göre 2 alt düğüme ayrılmıştır. Geleceğinden umutlu, evliliğinden orta derecede memnun ve kendine ayırdığı zamandan memnun ve çok memnun olan insanların(Düğüm 12) %59,2'si mutlu, geleceğinden umutlu, evliliğinden orta derecede memnun ve kendine ayırdığı zamandan memnun olmayan ve hiç memnun olmayan insanların(Düğüm 13) %41,9'unun mutlu olduğu sonucuna ulaşılmıştır.

Geleceğinden umutlu olmayanların oluşturduğu 2 nolu düğümü saflaştırmak için ise gelirinizden memnunmusunuz sorusu kullanılmış ve verilen yanıtlara göre gelirinizden orta derecede memnun olanlar, memnun olanlar ile çok memnun olanlar 7 nolu sol çocuk düğümünde, memnun olmayanlar ve hiç memnun olmayanlar 8 nolu sağ çocuk düğümde sınıflanmıştır. 6 ve 7 nolu düğümlerde saflaşma meydana geldiği için uç düğüm haline gelmiş ve sınıflamayı durdurmuşlardır.

CHAID algoritması ile oluşturulan ikinci sınıflama ağacı incelendiğinde mutluluğu etkileyen değişkenlerin gelecekte umutlu olmak, evlilikten, gelirinden ve kendine ayırdığı zamandan memnun olmak değişkenleri olduğu ve bu değişkenlerin arasında önemli bir ilişki olduğu sonucuna varılmıştır.

Mutluluğu etkileyen en önemli değişkenin gelecekte umutlu değişkeni olduğu sonucuna varılmıştır.Bu sonuca umut seviyesi düştükçe mutluluk oranları da önemli derecede azalmakta olduğu görülmektedir.

CHAID algoritması ile üçüncü model oluşturulurken veri setinin %70'i modeli oluşturmak, %30'u da modeli test etmek için kullanılmıştır.Bunun sonucunda oluşturulan ağaç CHAID-3 olarak adlandırılmış ve aşağıda gösterildiği şekildedir.



Şekil 18. CHAID Algoritması İle Oluşturulan 3. Sınıflama Ağacı

CHAID algoritması ile oluşturulan üçüncü sınıflama ağacı incelendiğinde, mutluluk düzeyleri üzerinde birinci derecede etkili bağımsız değişkenin gelecekte umutlu olma, ikinci derecede etkili bağımsız değişkenin gelirden memnun olma ve evliliğinden memnun olma, üçüncü derecede etkili bağımsız değişkenin gelirden memnun olma olduğu belirlenmiştir. Sınıflama ağacında mutluluk değişkeni aile düğümünü ayıran ilk ayıraç gelecekte umutlu olup olunmadığı sorusudur. Bu soruya umutluyum diyen kişiler 1. Düğümü, umutlu değilim diyen kişiler 2. düğümü, çok umutluyum diyen kişiler 3. Düğümü ve hiç umutlu değilim diyenler 4. düğümü oluşturmaktadır. Geleceğinden umutlu olanların %57,5'i, geleceğinden umudu olmayanların ise %30,8'i, gelecekte çok umutlu olanların % 45,6'sı ve gelecekte hiç umutlu olmayanların % 16,3'ü mutludur.

Gelecekte umutlu olanların oluşturduğu 1 nolu düğümü saflaştırmak için etki eden değişken katılımcılara sorulan evliliğinden memnunmusunuz sorusudur. Bu soruya verilen yanıtlara göre 1. Düğüm 3 alt düğümüne(Düğüm 5, 6 ve 7) ayrılmıştır. Düğüm 5'i gelecekte umutlu ve evliliğinden memnun kişiler, Düğüm 6'yı gelecekte umutlu ve evliliğinden çok memnun kişiler, Düğüm 7'yi gelecekte memnun, evliliğinden orta derecede memnun, memnun olmayan ve hiç memnun olmayan kişiler oluşturmaktadır.

Gelecekte umutlu ve evliliğinden memnun kişilerin oluşturduğu 5. Düğüm gelirden memnun olma durumuna göre üç alt düğümüne (Düğüm 10,11 ve 12) ayrılmıştır. Gelecekte umutlu ve hem evliliğinden hem de gelirinden memnun kişilerin %69'u mutludur.

Gelecekte memnun, evliliğinden orta derecede memnun, memnun olmayan ve hiç memnun olmayan kişilerin oluşturduğu 7. Düğüm gelirden memnun olma durumuna göre üç alt düğümüne (Düğüm 13, 14 ve 15) ayrılmıştır. Gelecekte umutlu ve hem evliliğinden hem de gelirinden memnun kişilerin %63,5'i mutludur.

Gelecekte umutlu olmayanların oluşturduğu 2. düğüm gelirinizden memnun olma durumuna göre 2 alt düğümüne(Düğüm 8 ve Düğüm 9) ayrılmıştır.Buna göre geleceğinden umutlu olmayıp gelirinden orta derecede memnun, memnun, çok memnun olanların(Düğüm 8) %38,8'i mutlu, gelecekte umutlu olmayıp gelirinden memnun olmayanlar ile hiç memnun olmayanların(Düğüm 9) ise %24,1'i mutludur.

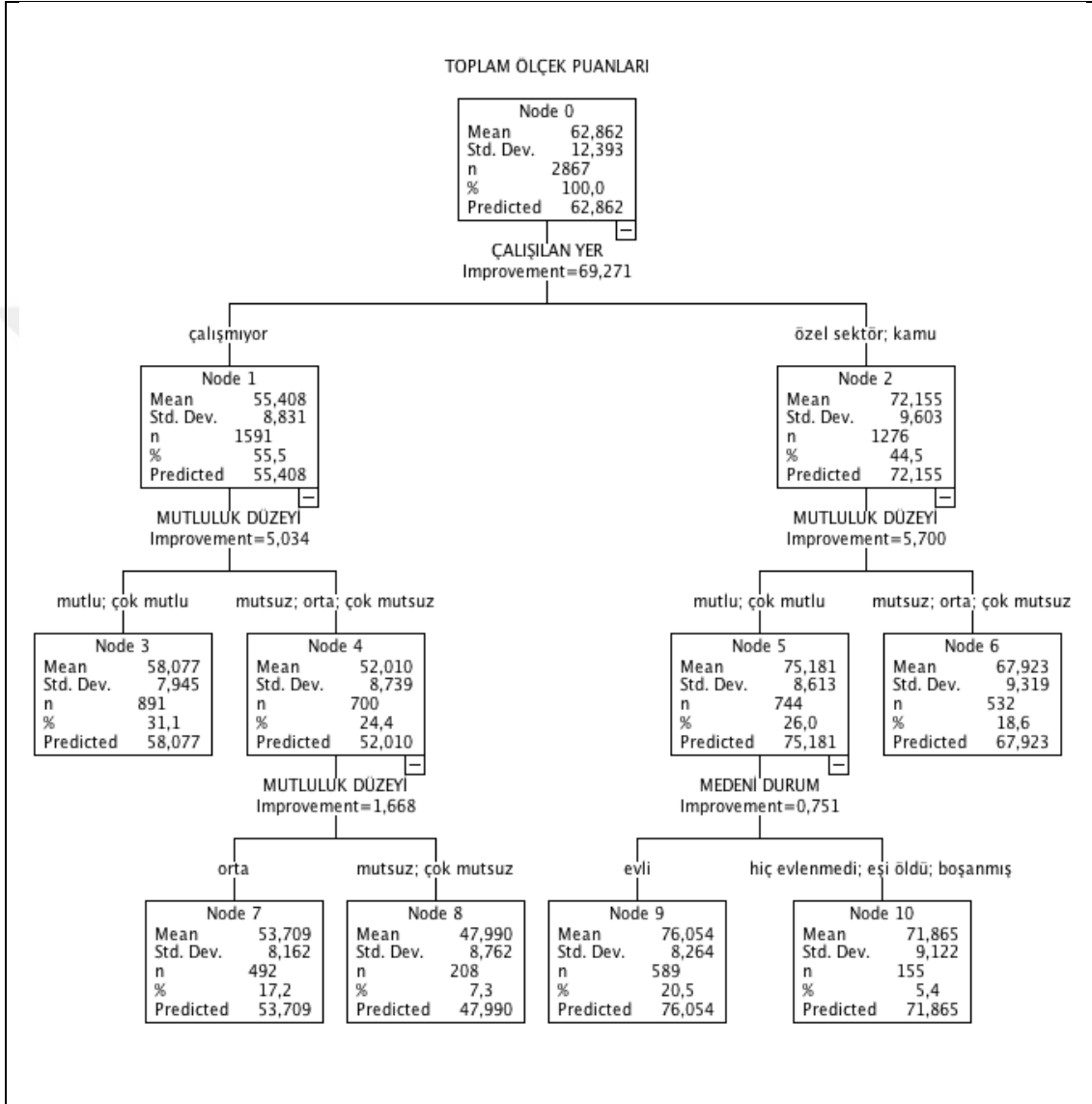
CHAID-3 ağacı incelendiğinde mutluluğu etkileyen değişkenlerin gelecekte umutlu olmak, evlilikten, gelirinden memnun olmak değişkenleri olduğu ve bu değişkenlerin arasında önemli bir ilişki olduğu sonucuna varılmıştır.

Mutluluğu etkileyen en önemli değişkenin gelecekte umutlu değişkeni olduğu sonucuna varılmıştır.Umut seviyesi düştükçe mutluluk oranları da önemli derecede azalmaktadır. Gelecekte umutlu olanları evlilik, umutlu olmayanları gelirden memnun olma etkilemektedir. Evlilikten memnun olma durumunu etkileyen gelir değişkeni sonuçlarına bakıldığı zaman, geleceğinden umutlu, evliliğinden memnun fakat gelirinden memnun olmayan mutlu kişilerin oranının, geleceğinden umutlu, evliliğinden ve gelirinden memnun olmayan mutlu kişilerin orandan yüksek olduğu sonucuna ulaşılmıştır.

Sonuç olarak bütün sınıflama ağaçları incelendiğinde mutluluğu etkileyen en önemli değişkenin gelecekte umutlu olma değişkeni olduğu belirlenmiştir. Model veri oranı arttıkça mutluluğu etkileyen diğer faktörlerin; evliliğinden, gelirinden ve kendine ayırdığı zamandan memnun olma değişkenleri olduğu görülmüştür. Yani; model verisi arttıkça mutluluğu etkileyen değişken sayısı ve düğüm sayısı da artmaktadır ve ağaç daha karmaşık bir yapıya bürünmektedir. Dolayısıyla, kategoriler yani sınıflar hakkında daha ayrıntılı bilgiye ulaşılmaktadır. Sınıflar arasındaki etkileşimlerde görülebilmektedir. Bu durumda da araştırmacı analiz ettiği veriye ait daha detaylı bilgiye sahip olacağından model verisinin yüksek tutulması önerilmektedir.

3.2.4 CART Algoritması İle Oluşturulan Regresyon Ağaçlarına Ait Bulgular

İlk olarak veri setinin %30'u modeli oluşturmak, %70'i de modeli test etmek için kullanılmıştır. Oluşan model Şekil 19'da gösterilmektedir.



Şekil 19. CART Algoritması İle Oluşturulan 1. Regresyon Ağacı

Şekil 19 incelendiğinde memnuniyet puan ortalaması (kök düğüm) yaklaşık 62,9 (S = 12,4) olarak tahmin edilmiştir. Memnuniyet puanını birinci derecede etkileyen değişken çalışılan kurum olarak bulunmuştur. Memnuniyet puanını daha sonra sırasıyla mutluluk düzeyi ve medeni durum etkilemektedir.

Düğüm 1 olarak adlandırılan birinci alt grupta; çalışmayanlar olup, çalışmayanlara ait ortalama memnuniyet puanı 55,41 olarak bulunmuştur. Bu düğüm toplam veri adedinin % 55,5'ini oluşturmuştur. Düğüm 2, yani ikinci alt grup; özel sektörde

çalışanlar ile kamuda çalışanlardan oluşmakta ve bunların ortalama memnuniyet puanları 72,155 olarak bulunmuştur. Bu düğüm ise toplam veri adedinin % 44,5'ini oluşturmuştur. Düğüm 1 ve Düğüm 2'ye bakıldığında çalışan kişilerin memnuniyet puanlarının daha yüksek olduğu sonucuna varılmıştır.

Düğüm 1 ve Düğüm 2, Mutluluk Düzeyi değişkenine göre 2'şer alt gruba (Düğüm 3, 4, 5, 6) ayrılmıştır. Düğüm 3'ü oluşturan alt grup; yani, Düğüm 1'de bulunan çalışmayan kişilerden mutlu ve çok mutlu olanların ortalama memnuniyet puanları 58,1 olarak bulunmuştur. Düğüm 1'de bulunan çalışmayan kişilerden mutluluk düzeyi; orta, mutsuz ve çok mutsuz olanlar 4. Düğümü oluşturmaktadır. Mutluluk düzeyi orta, mutsuz ve çok mutsuz olan çalışmayan kişilerin memnuniyet puanlarının ortaması ise 52'dir. Düğüm 3 ve Düğüm 4'e bakıldığı zaman çalışmayan kişilerde kendini mutlu ve çok mutlu hisseden kişilerin kendini orta, mutsuz ve çok mutsuz hissedenlere göre memnuniyet puanlarının daha yüksek olduğu sonucuna ulaşılmıştır.

Mutluluk düzeyi orta, mutsuz ve çok mutsuz olan çalışmayan kişilerin oluşturduğu 4. Düğüm tekrardan mutluluk düzeyine göre 2 alt düğüme (Düğüm 7 ve 8) ayrılmıştır. Düğüm 7'de çalışmayan ve mutluluk düzeyi orta seviyede olan kişilerin memnuniyet puanlarının ortalaması 53,7 ; Düğüm 8'de bulunan çalışmayan ve kendini mutsuz ve çok mutsuz hisseden kişilerin memnuniyet puan ortalaması ise 47,9 olarak bulunmuştur. Bu sonuca göre, çalışmayan ve kendini orta seviyede mutlu hisseden kişilerin çalışmayan kendini mutsuz ve çok mutsuz hisseden kişilere göre memnuniyet puanlarının biraz daha yüksek olduğu görülmüştür.

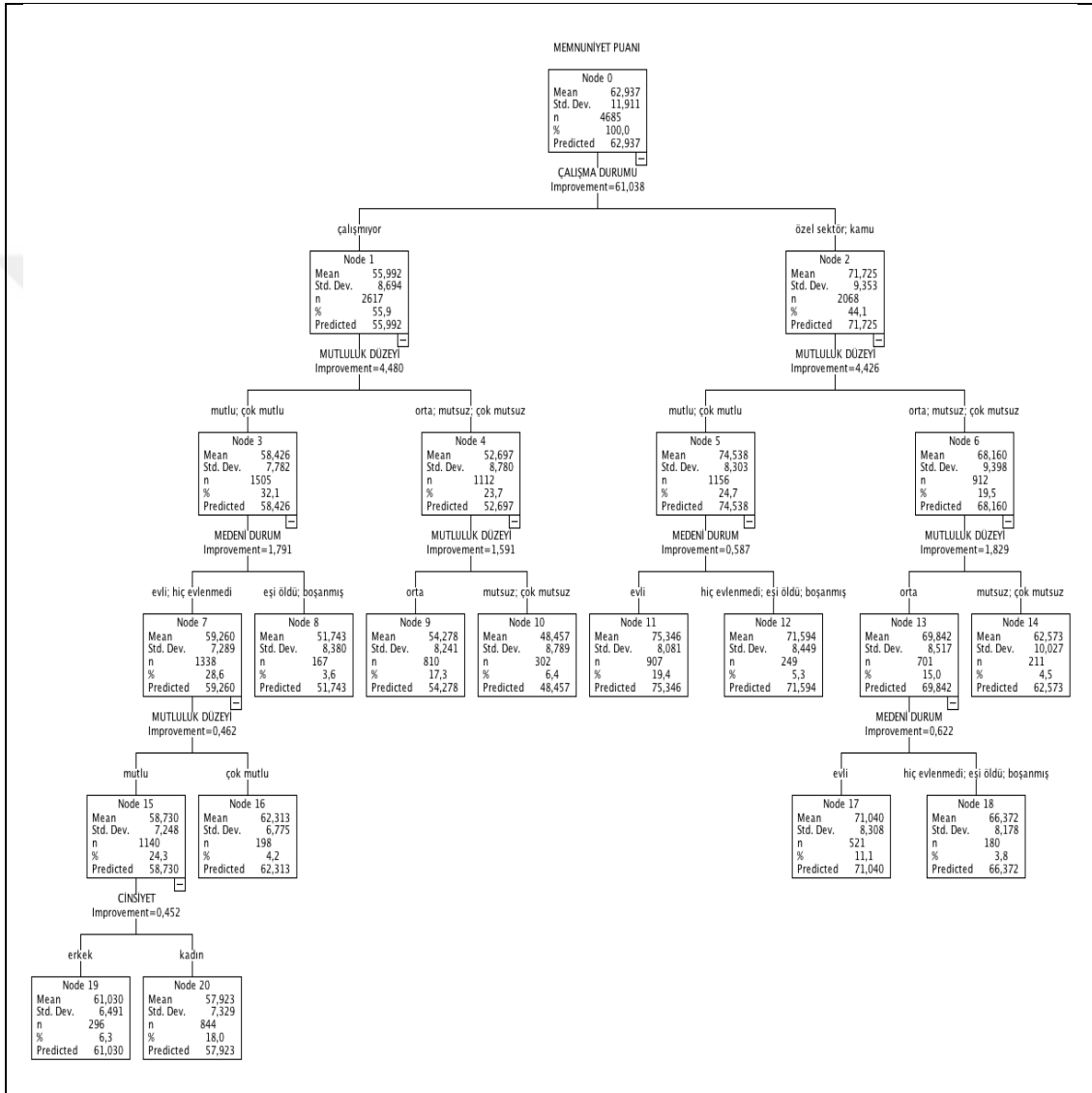
Düğüm 2'de bulunan özel sektörde çalışan ve kamu sektöründe çalışan kişilerden mutluluk düzeyi çok mutlu ve mutlu olanlar Düğüm 5'i, mutluluk düzeyi; orta, mutsuz ve çok mutsuz olanlar 6. Düğümü oluşturmaktadır.

Mutluluk düzeyi çok mutlu ve mutlu olan özel veya kamu sektöründe çalışan kişilerin memnuniyet puanlarının ortalaması 75,2 ; mutluluk düzeyi orta, mutsuz ve çok mutsuz olan özel veya kamu sektöründe çalışan kişilerin memnuniyet puanlarının ortaması ise 67,9'dur. Düğüm 5 ve Düğüm 6'ya bakıldığı zaman özel sektörde veya kamu sektöründe çalışan kişilerde kendini mutlu ve çok mutlu hisseden kişilerin kendini orta, mutsuz ve çok mutsuz hissedenlere göre memnuniyet puanlarının daha yüksek olduğu sonucuna ulaşılmıştır. Özel sektörde veya kamu sektöründe çalışıp kendini mutlu ve çok mutlu hisseden kişilerin oluşturduğu 4. Düğüm, Medeni Durum değişkenine göre 2 alt düğüme (Düğüm 9 ve 10) ayrılmıştır. Düğüm 9'da özel veya kamu sektöründe çalışan ve mutluluk düzeyi mutlu ve çok mutlu olan evli kişilerin memnuniyet puanlarının ortalaması 76 ; Düğüm 8'de bulunan özel veya kamu sektöründe çalışan ve kendini mutlu ve çok mutlu hisseden hiç evlenmemiş, eşi ölmüş veya boşanmış kişilerin memnuniyet puan ortalaması ise 71,8 olarak bulunmuştur.

CART algoritması ile oluşturulan ilk regresyon ağacına bakıldığı zaman kişilerin memnuniyet puanlarını etkileyen en önemli değişkenin çalışma durumu olduğu görülmüştür. Özel yada kamuda çalışanların çalışmayanlara göre memnuniyet puanları daha yüksektir. Memnuniyet puanlarını çalışma durumundan sonra

mutluluk düzeyi ve medeni durumları etkilemektedir. Özel yada kamuda çalışan kendini mutlu hisseden evli kişilerin memnuniyet puanları diğer medeni durumdaki kişilere göre oldukça yüksektir.

CART algoritmasıyla oluşturulmuş ikinci regresyon ağacında veri setinin yarısı modeli test etmek diğer yarısı modeli oluşturmak için kullanılmıştır. Veri setinin yarısı test etmek için, yarısı model oluşturmak için kullanıldığı zaman oluşan ağaç şekil 20’de gösterilmiştir.



Şekil 20. CART Algoritması İle Oluşturulan 2. Regresyon Ağacı

Şekil 20 incelendiğinde memnuniyet puan ortalaması (kök düğüm) yaklaşık 62,92 (S = 11,9) olarak tahmin edilmiştir. Memnuniyet puanını birinci derecede etkileyen değişken çalışılan kurum olarak bulunmuştur. Memnuniyet puanını daha sonra ise mutluluk düzeyi, medeni durum ve cinsiyet etkilemektedir.

Düğüm 1 olarak adlandırılan birinci alt grupta; çalışmayanlar olup, çalışmayanlara ait ortalama memnuniyet puanı yaklaşık 56 olarak bulunmuştur. Bu düğüm toplam veri adedinin % 55,9’sini oluşturmuştur. Düğüm 2, yani ikinci alt grup; özel sektörde

çalışanlar ile kamuda çalışanlardan oluşmakta ve bunların ortalama memnuniyet puanları 71,72 olarak bulunmuştur. Bu düğüm ise toplam veri adedinin % 44,1'ünü oluşturmuştur. Düğüm 1 ve Düğüm 2'ye bakıldığında çalışan kişilerin memnuniyet puanlarının daha yüksek olduğu sonucuna varılmıştır.

Düğüm 1 ve Düğüm 2, Mutluluk düzeyi değişkenine göre 2'şer alt gruba (Düğüm 3, 4, 5, 6) ayrılmıştır. Düğüm 3'ü oluşturan alt grup; yani, Düğüm 1'de bulunan çalışmayan kişilerden mutlu ve çok mutlu olanların ortalama memnuniyet puanları 58,43 olarak bulunmuştur.

Düğüm 1'de bulunan çalışmayan kişilerden mutluluk düzeyi; orta, mutsuz ve çok mutsuz olanlar 4. Düğümü oluşturmaktadır. Mutluluk düzeyi orta, mutsuz ve çok mutsuz olan çalışmayan kişilerin memnuniyet puanlarının ortaması ise 52,7'dir. Düğüm 3 ve Düğüm 4'e bakıldığı zaman çalışmayan kişilerde kendini mutlu ve çok mutlu hisseden kişilerin kendini orta, mutsuz ve çok mutsuz hissedenlere göre memnuniyet puanlarının daha yüksek olduğu sonucuna ulaşılmıştır.

Mutluluk düzeyi mutlu ve çok mutlu olan çalışmayan kişilerin oluşturduğu 3. Düğüm tekrardan medeni duruma göre 2 alt düğüme (Düğüm 7 ve Düğüm 8) ayrılmıştır. Düğüm 7 de mutluluk düzeyi mutlu ve çok mutlu olan çalışmayan evli ve hiç evlenmemiş kişilerin memnuniyet puanı 59,3 ; Düğüm 8 de mutluluk düzeyi mutlu ve çok mutlu olan çalışmayan eşi ölmüş ve boşanmış kişilerin memnuniyet puanları 51,7 bulunmuştur.

Mutluluk düzeyi mutlu ve çok mutlu olan evli ve hiç evlenmemiş kişiler mutluluk düzeyine göre (Düğüm 15 ve Düğüm 16) tekrardan 2 alt düğüme ayrılmıştır. Düğüm 15 te mutluluk düzeyi mutlu olan evli ve hiç evlenmemiş kişiler cinsiyetlerine göre (Düğüm 19 ve Düğüm 20) 2 alt düğüme ayrılmıştır. Ağaç modeline göre evli ve hiç evlenmemiş mutlu erkeklerin memnuniyet puanları(61,0), evli ve hiç evlenmemiş mutlu kadınların memnuniyet puanlarından(57,9) biraz daha yüksek olduğu görülmüştür.

Mutluluk düzeyi orta, mutsuz ve çok mutsuz olan çalışmayan kişilerin oluşturduğu 4. Düğüm tekrardan mutluluk düzeyine göre 2 alt düğüme (Düğüm 9 ve 10) ayrılmıştır. Düğüm 9'da çalışmayan ve mutluluk düzeyi orta seviyede olan kişilerin memnuniyet puanlarının ortalaması 54,3 ; Düğüm 10'da bulunan çalışmayan ve kendini mutsuz ve çok mutsuz hisseden kişilerin memnuniyet puan ortalaması ise 48,6 olarak bulunmuştur. Bu sonuca göre, çalışmayan ve kendini orta seviyede mutlu hisseden kişilerin çalışmayan kendini mutsuz ve çok mutsuz hisseden kişilere göre memnuniyet puanlarının biraz daha yüksek olduğu görülmüştür.

Düğüm 2'de bulunan özel sektörde çalışan ve kamu sektöründe çalışan kişilerden mutluluk düzeyi çok mutlu ve mutlu olanlar Düğüm 5'i, mutluluk düzeyi; orta, mutsuz ve çok mutsuz olanlar 6. Düğümü oluşturmaktadır.

Mutluluk düzeyi çok mutlu ve mutlu olan özel veya kamu sektöründe çalışan kişilerin memnuniyet puanlarının ortalaması 74,5 ; mutluluk düzeyi orta, mutsuz ve çok mutsuz olan özel veya kamu sektöründe çalışan kişilerin memnuniyet puanlarının ortaması ise 68,2'dir. Düğüm 5 ve Düğüm 6'ya bakıldığı zaman özel sektörde veya kamu sektöründe çalışan kişilerde kendini mutlu ve çok mutlu hisseden kişilerin kendini orta, mutsuz ve çok mutsuz hissedenlere göre memnuniyet puanlarının daha yüksek olduğu sonucuna ulaşılmıştır.

Düğüm 5' i oluşturan özel veya kamu sektöründe çalışan kendini mutlu ve çok mutlu hissedenler medeni durumlarına göre 2 alt düğüme (Düğüm 11 ve Düğüm 12)

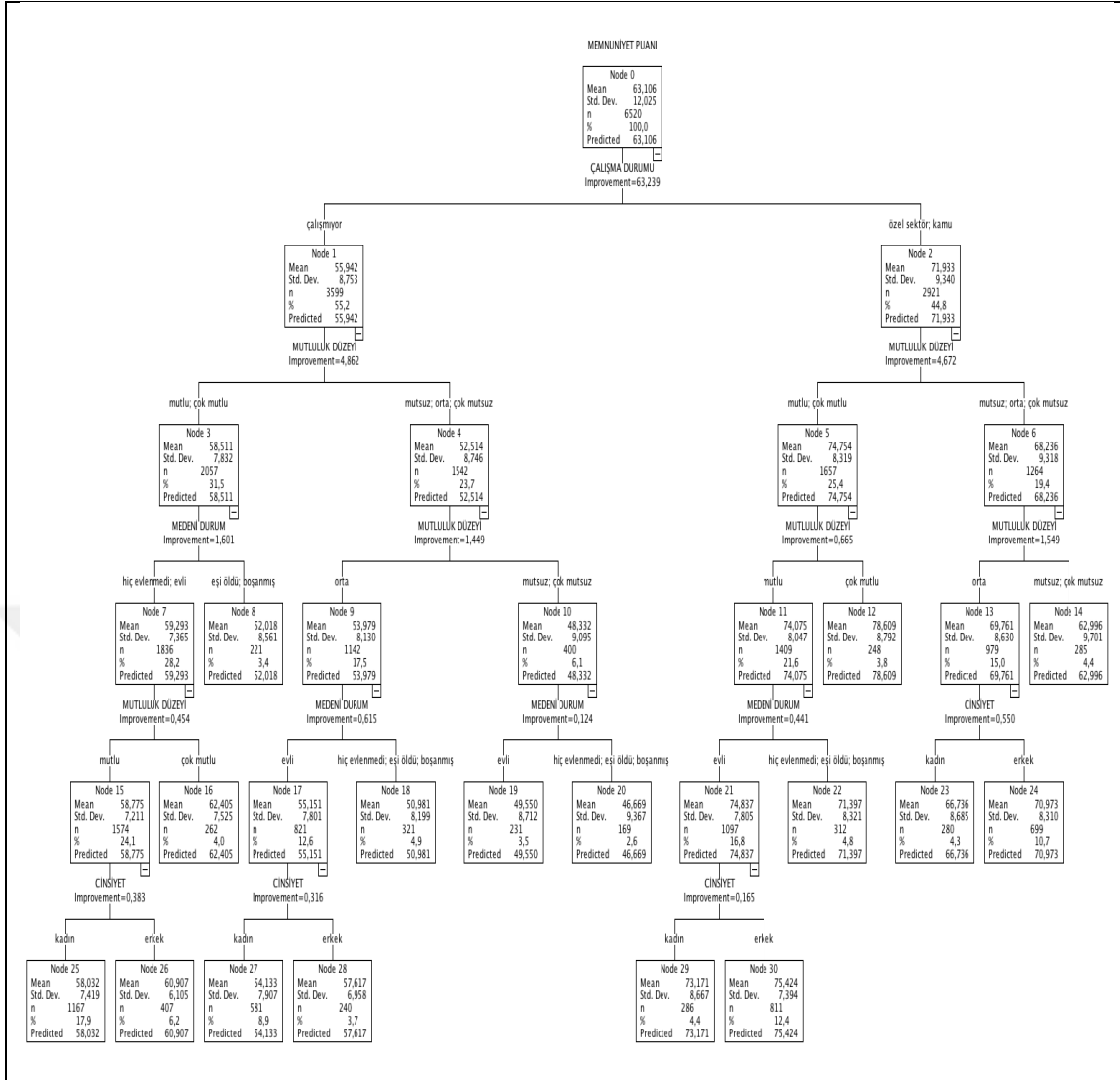
ayrılmıştır. Dügüm 11'i oluşturan özel veya kamu sektöründe çalışan kendini mutlu ve çok mutlu hisseden evli kişilerin memnuniyet puanlarının (75,3), Dügüm 12'yi oluşturan özel veya kamu sektöründe çalışan kendini mutlu ve çok mutlu hisseden hiç evlenmemiş, eşi ölmüş yada ayrılmış kişilerin memnuniyet puanlarından (71,6) biraz daha yüksek olduğu görülmüştür.

Dügüm 6'yı oluşturan mutluluk düzeyi; orta, mutsuz ve çok mutsuz olan özel veya kamuda çalışan kişiler mutluluk düzeyine göre tekrardan iki alt düğüme (Dügüm 13 ve Dügüm 14) ayrılmıştır.

Dügüm 13'ü oluşturan mutluluk düzeyi orta olan özel veya kamuda çalışan kişiler ise medeni durumlarına 2 alt düğüme (Dügüm 17 ve Dügüm 18) ayrılmıştır. Dügüm 17'de mutluluk düzeyi orta olan özel veya kamuda çalışan evli kişilerin memnuniyet puanlarının (71,0), mutluluk düzeyi orta olan özel veya kamuda çalışan evlenmemiş, eşi ölmüş yada boşanmış kişilerin memnuniyet puanlarından (66,4) biraz daha yüksek olduğu görülmüştür.

CART algoritması ile oluşturulan ikinci regresyon ağacına bakıldığı zaman kişilerin memnuniyet puanlarını etkileyen en önemli değişkenin çalışma durumu olduğu görülmüştür. Özel yada kamuda çalışanların çalışmayanlara göre memnuniyet puanları daha yüksektir. Memnuniyet puanlarını çalışma durumundan sonra mutluluk düzeyi ve medeni durumları ve cinsiyet etkilemektedir. Özel yada kamuda çalışan kendini mutlu hisseden evli kişilerin memnuniyet puanları diğer medeni durumdaki kişilere göre oldukça yüksektir. Evli ve hiç evlenmemiş mutlu erkeklerin memnuniyet puanları ise evli ve hiç evlenmemiş mutlu kadınların memnuniyet puanlarından biraz daha yüksek olduğu görülmüştür

CART algoritmasıyla oluşturulmuş üçüncü regresyon ağacında veri setinin %70'i modeli oluşturmak %30'u modeli test etmek için kullanılmıştır. Veri setinin %70'i modeli oluşturmak, %30'u test etmek için kullanıldığı zaman oluşan ağaç CART-6 ağacı olarak adlandırılmış ve şekil 21'de gösterilmiştir.



Şekil 21. CART Algoritması İle Oluşturulan 3. Regreyon Ağacı

Şekil 21 incelendiğinde memnuniyet puan ortalaması (kök düğüm) yaklaşık 63,1 ($S = 12,02$) olarak tahmin edilmiştir. Memnuniyet puanını birinci derecede etkileyen değişken çalışılan kurum olarak bulunmuştur. Memnuniyet puanını daha sonra ise mutluluk düzeyi, medeni durum ve cinsiyet değişkenleri etkilemektedir.

Düğüm 1 olarak adlandırılan birinci alt grupta; çalışmayanlar olup, çalışmayanlara ait ortalama memnuniyet puanı 55,8 olarak bulunmuştur. Bu düğüm toplam veri adedinin % 55,7'sini oluşturmaktadır. Düğüm 2, yani ikinci alt grup; özel sektörde çalışanlar ile kamuda çalışanlardan oluşmakta ve bunların ortalama memnuniyet puanları 72,84 olarak bulunmuştur. Bu düğüm ise toplam veri adedinin % 44,3'ünü oluşturmaktadır. Düğüm 1 ve Düğüm 2'ye bakıldığında çalışan kişilerin memnuniyet puanlarının daha yüksek olduğu sonucuna varılmıştır.

Düğüm 1 ve Düğüm 2, Mutluluk Düzeyi değişkenine göre 2'şer alt gruba (Düğüm 3, 4, 5, 6) ayrılmıştır. Düğüm 3'ü oluşturan alt grup; yani, Düğüm 1'de bulunan çalışmayan kişilerden mutlu ve çok mutlu olanların ortalama memnuniyet puanı 58,3 olarak bulunmuştur.

Düğüm 1'de bulunan çalışmayan kişilerden mutluluk düzeyi; orta, mutsuz ve çok mutsuz olanlar 4. Düğümü oluşturmaktadır. Mutluluk düzeyi orta, mutsuz ve çok

mutsuz olan çalışmayan kişilerin memnuniyet puan ortaması ise 52,6'dır. Düğüm 3 ve Düğüm 4'e bakıldığı zaman çalışmayan kişilerde kendini mutlu ve çok mutlu hisseden kişilerin kendini orta, mutsuz ve çok mutsuz hissedenlere göre memnuniyet puanlarının daha yüksek olduğu sonucuna ulaşılmıştır.

Düğüm 2'de bulunan özel sektörde çalışan ve kamu sektöründe çalışan kişilerden mutluluk düzeyi çok mutlu ve mutlu olanlar Düğüm 5'i, mutluluk düzeyi; orta, mutsuz ve çok mutsuz olanlar 6. Düğümü oluşturmaktadır.

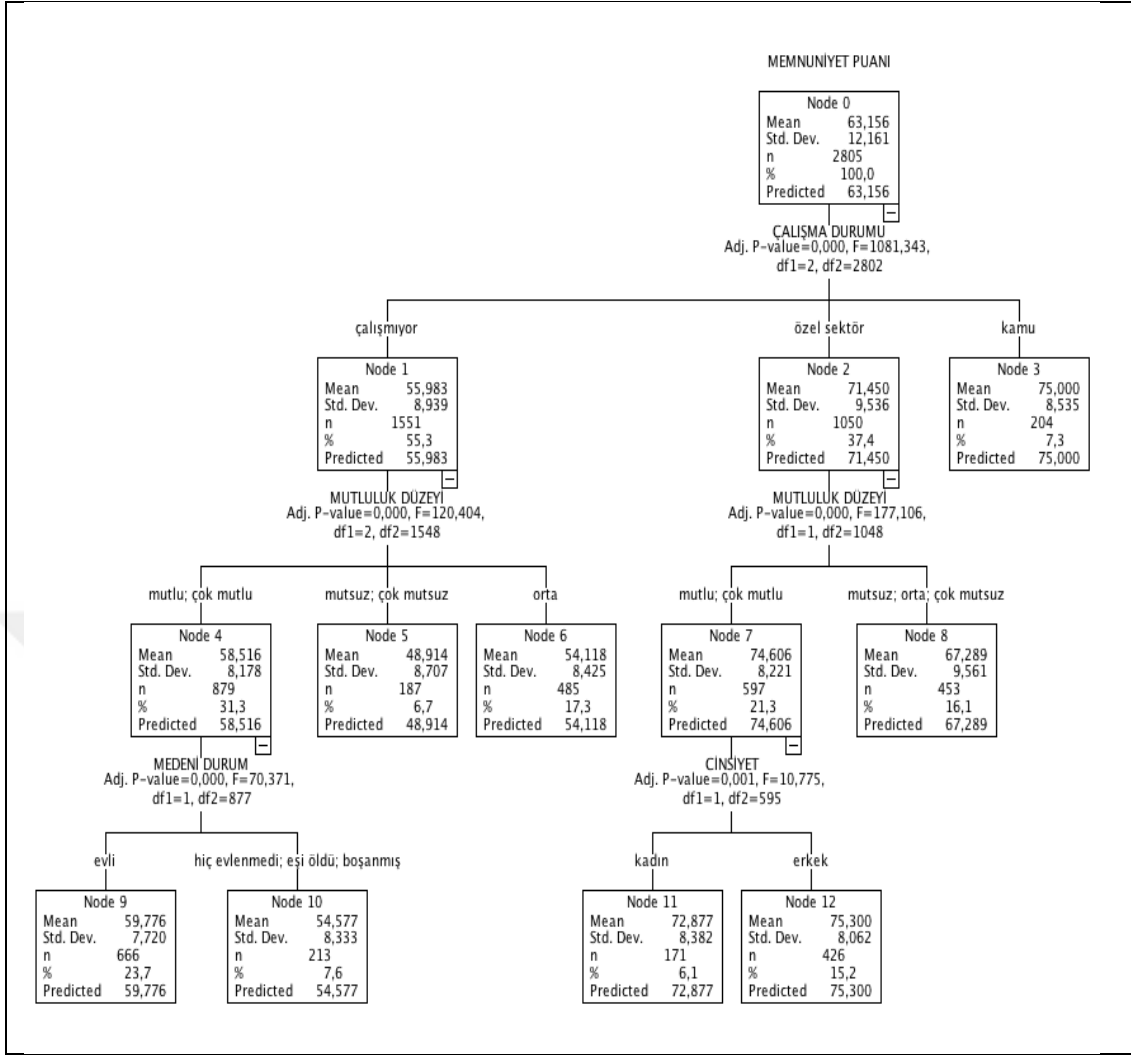
Mutluluk düzeyi çok mutlu ve mutlu olan özel veya kamu sektöründe çalışan kişilerin memnuniyet puanlarının ortalaması 74,8; mutluluk düzeyi orta, mutsuz ve çok mutsuz olan özel veya kamu sektöründe çalışan kişilerin memnuniyet puanlarının ortaması ise 68,1'dir. Düğüm 5 ve Düğüm 6'ya bakıldığı zaman özel sektörde veya kamu sektöründe çalışan kişilerde kendini mutlu ve çok mutlu hisseden kişilerin kendini orta, mutsuz ve çok mutsuz hissedenlere göre memnuniyet puanlarının daha yüksek olduğu sonucuna ulaşılmıştır.

CART algoritması ile oluşturulan regresyon ağaçları incelendiğinde, memnuniyet puanlarını etkileyen en önemli değişkenin kişilerin çalışma durumları olduğu belirlenmiştir. Oluşturulan modeller sonucunda özel sektörde ya da kamuda çalışan kişilerin mutluluk düzeyleri ne olursa olsun genel olarak çalışmayanlara göre memnuniyet puanlarının daha yüksek, yani hayatlarından daha memnun olduğu sonucuna ulaşılmıştır. Mutluluk düzeylerine bakıldığında çalışan kişilerde de çalışmayan kişilerde de mutlu ve çok mutlu olanların memnuniyet puanları, mutluluk düzeyi orta, mutsuz, çok mutsuz olanlara göre daha yüksektir. Model veri oranı arttıkça memnuniyet puanlarını; mutluluk düzeyleri, medeni durum ve cinsiyet değişkenlerinin etkilediği sonucuna ulaşılmıştır. Kişiyi kimin mutlu ettiği ve neyin mutlu ettiği kişilerin memnuniyet puanını etkilemediği görülmüştür. Ayrıca model veri oranı artsa da, memnuniyet puanını 1. dereceden ve 2. dereceden etkileyen değişkenler farklılık göstermediği sonucuna ulaşılmıştır. Ancak ağaçlar arasındaki en önemli farklılık model verisi oranı arttıkça bağımlı değişkeni etkileyen bağımsız değişken sayısının ve düğüm sayılarının artmasıdır. Dolayısıyla, kategoriler yani sınıflar hakkında daha ayrıntılı bilgiye ulaşılmaktadır. Sınıflar arasındaki etkileşimlerde görülebilmektedir. Bu durumda, ağaç belirlenirken model verisinin en yüksek olduğu ağacın tercih edilmesi önerilmektedir.

3.2.5 CHAID Algoritması İle Oluşturulan Regresyon Ağaçlarına Ait Bulgular

CHAID algoritması ile oluşturulan regresyon ağacında, bağımlı değişkeni en çok etkileyen bağımsız değişken en yüksek F değerine sahip olmaktadır.

CHAID algoritması ile model oluşturulurken ilk olarak veri setinin %30'u modeli oluşturmak, %70'i de modeli test etmek için kullanılmıştır. Bunun sonucunda oluşturulan ağaç şekil 22'de gösterildiği gibidir.



Şekil 22. CHAID Algoritması İle Oluşturulan 1. Regresyon Ağacı

CHAID algoritması ile oluşturulan ilk regresyon ağacı incelendiğinde, memnuniyet puanları üzerinde birinci derecede etkili bağımsız değişkenin çalışılan yer (Düz. P=0.000), ikinci derecede etkili bağımsız değişkenin mutluluk düzeyi (Düz. P=0.000), üçüncü derecede etkili bağımsız değişkenlerin ise medeni durum (Düz. P=0.000) ve cinsiyet (Düz. P=0.001) olduğu belirlenmiştir.

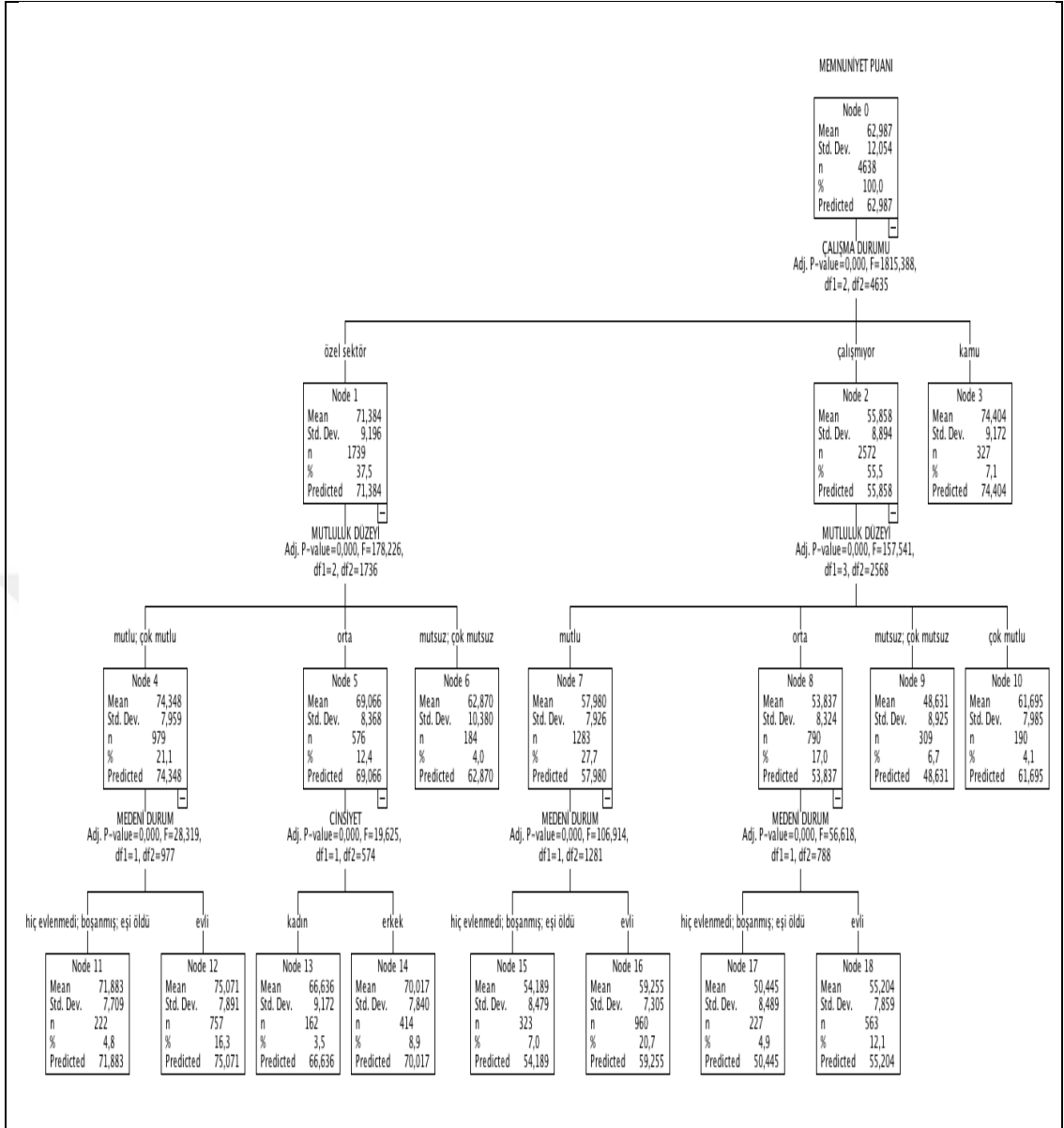
Memnuniyet puanları (Düğüm 0), çalışılan yer bakımından 3 alt gruba (Düğüm 1, Düğüm 2, Düğüm 3) ayrılmıştır. Düğüm 1; çalışmayanların oluşturduğu alt grubu, Düğüm 2; özel sektörde çalışanların oluşturduğu alt grubu, Düğüm 3; kamuda çalışanların oluşturduğu alt grubu temsil etmektedir. Düğüm 1 (çalışmayanların) memnuniyet puanı ortalaması 55,98 (S=8,93), Düğüm 2 (özel sektörde çalışanların) memnuniyet puan ortalaması 71,45 (S=9,5), Düğüm 3 (kamuda çalışanların) memnuniyet puan ortalaması 75 (S=8,53) olarak bulunmuştur. Oluşturulan alt gruplardan yeterince homojen yapıya sahip olan Düğüm 3'e terminal düğüm adı verilir. Ayrıca memnuniyet puan ortalamalarına bakıldığında memnuniyet puanları sırasıyla kamuda çalışanlar, özel sektörde çalışanlar ve çalışmayanlar olarak belirlenmiştir.

Düğüm 1’de çalışmayanların memnuniyet puanları, mutluluk düzeyi değişkeni tarafından etkilenmiştir (Düz P=0.000). Düğüm 1 (çalışmayan kişiler) mutluluk düzeyi bakımından üç yeni alt gruba (Düğüm 4,5 ve 6) ayrılmıştır. Düğüm 4; mutluluk düzeyi mutlu ve çok mutlu olan çalışmayanların oluşturduğu alt grubu, Düğüm 5; mutluluk düzeyi mutsuz ve çok mutsuz olan çalışmayanların oluşturduğu alt grubu, Düğüm 6; mutluluk düzeyi orta olan çalışmayanların ifade etmektedir. Düğüm 4, Düğüm 5 ve Düğüm 6 için memnuniyet puan ortalamaları sırasıyla, 58,5 (S=8,2) , 48,9 (S=8,7) ve 54.1 (S=8.4) olarak tahmin edilmiştir. Mutluluk düzeyi mutlu ve çok mutlu olan çalışmayan kişilerin (Düğüm 4) memnuniyet puanlarının medeni durum tarafından etkilendiği belirlenmiştir. Bu sonuca göre mutluluk düzeyi mutlu ve çok mutlu olan evli kişilerin diğer medeni durumdakilere göre memnuniyet puanunun daha yüksek olduğu belirlenmiştir.

Düğüm 2’de özel sektörde çalışan kişilerin memnuniyet puanları analize dahil edilen bağımsız değişkenlerden sadece mutluluk düzeyi tarafından etkilendiği görülmüştür (Düz. P=0.000). Düğüm 3; mutluluk düzeyi bakımından iki yeni alt gruba (Düğüm 7 ve Düğüm 8) ayrılmıştır. Düğüm 7; özel sektörde çalışan ve mutluluk düzeyi mutlu ve çok mutlu olan kişilerin oluşturduğu alt grubu, Düğüm 8; özel sektörde çalışan ve mutluluk düzeyleri orta, mutsuz, çok mutsuz kişilerin oluşturduğu alt grubu ifade etmektedir. Düğüm 7 ve Düğüm 8 için memnuniyet puan ortalamaları sırasıyla, 74,6 (S=8,2) ve 67,3 (S=9,6) olarak hesaplanmıştır. Mutluluk düzeyi mutlu ve çok mutlu olan ve özel sektörde çalışan kişilerin memnuniyet puanlarının sadece cinsiyet tarafından etkilendiği saptanmıştır (Düz. P= 0.012). Bu sonuca göre mutluluk düzeyi mutlu ve çok mutlu olan ve özel sektörde çalışan erkeklerin kadınlara göre memnuniyet puanları daha yüksek olduğu görülmüştür.

CHAID algoritması ile oluşturulan ilk regresyon ağacına bakıldığı zaman kişilerin memnuniyet puanlarını etkileyen en önemli değişkenin çalışma durumu olduğu görülmüştür. Kamuda çalışanların memnuniyet puanları özel sektörde çalışanlardan, özel sektörde çalışanların çalışmayanlara göre memnuniyet puanları daha yüksektir. Memnuniyet puanlarını çalışma durumundan sonra mutluluk düzeyi ve medeni durumları ve cinsiyet etkilemektedir. Çalışmayan ve mutlu olan evli kişilerin diğer medeni durumdakilere göre memnuniyet puanunun daha yüksek olduğu belirlenmiştir. Özel sektörde çalışan ve mutlu olan erkeklerin ise kadınlara göre memnuniyet puanları daha yüksek olduğu görülmüştür.

CHAID algoritması ile ikinci regresyon modeli oluşturulurken veri setinin yarısı modeli oluşturmak, yarısı da modeli test etmek için kullanılmıştır. Bunun sonucunda oluşturulan ağaç şekil 23’te gösterildiği şekildedir.



Şekil 23. CHAID Algoritması İle Oluşturulan 2. Regresyon Ağacı

CHAID algoritması ile oluşturulan ikinci regresyon ağacı incelendiğinde, memnuniyet puanları üzerinde birinci derecede etkili bağımsız değişkenin çalışılan yer (Düz. P=0.000), ikinci derecede etkili bağımsız değişkenin mutluluk düzeyi (Düz. P=0.000), üçüncü derecede etkili bağımsız değişkenin medeni durum (Düz. P=0.000) ve cinsiyet (Düz. P=0.000) olduğu belirlenmiştir.

Memnuniyet puanları (Düğüm 0), çalışılan yer bakımından 3 alt gruba (Düğüm 1, Düğüm 2, Düğüm 3) ayrılmıştır. Düğüm 1; özel sektörde çalışanların oluşturduğu alt grubu, Düğüm 2; çalışmayanların oluşturduğu alt grubu, Düğüm 3; kamuda çalışanların oluşturduğu alt grubu temsil etmektedir. Düğüm 1 (özel sektörde çalışanların) memnuniyet puanı ortalaması 71,4 (S=9,2), Düğüm 2 (çalışmayanların) memnuniyet puan ortalaması 55,9 (S=8,9), Düğüm 3 (kamuda çalışanların) memnuniyet puan ortalaması 74,4 (S=9,2) olarak bulunmuştur. Oluşturulan alt

gruplardan yeterince homojen yapıya sahip olan Dügüm 3'e terminal düğüm adı verilir. Ayrıca memnuniyet puan ortalamalarına bakıldığında memnuniyet puanları sırasıyla kamuda çalışanlar, özel sektörde çalışanlar ve çalışmayanlar olarak belirlenmiştir.

Dügüm 1'de özel sektörde çalışan kişilerin memnuniyet puanları analize dahil edilen bağımsız değişkenlerden mutluluk düzeyi tarafından etkilenmektedir (Düz. $P=0.000$). Dügüm 1; mutluluk düzeyi bakımından üç yeni alt gruba (Dügüm 4,5 ve 6) ayrılmıştır. Dügüm 4; özel sektörde çalışan ve mutluluk düzeyi mutlu ve çok mutlu olan kişilerin oluşturduğu alt grubu, Dügüm 5; özel sektörde çalışan ve mutluluk düzeyleri orta olan kişilerin oluşturduğu alt grubu, Dügüm 6; özel sektörde çalışan mutluluk düzeyi mutsuz, çok mutsuz olan kişilerin oluşturduğu alt grubu ifade etmektedir. Dügüm 4, Dügüm 5 ve Dügüm 6 için memnuniyet puan ortalamaları sırasıyla, 74,3 ($S=7,96$), 69,1 ($S=8,4$) ve 62,9 ($S=10,4$) olarak hesaplanmıştır.

Mutluluk düzeyi, mutlu ve çok mutlu olan özel sektörde çalışanların (Dügüm 4) memnuniyet puanları sadece medeni durum tarafından etkilenmektedir. (Düz. $P=0.000$). Bu sonuca göre mutluluk düzeyi, mutlu ve çok mutlu olan ve özel sektörde çalışan kişilerde evlilerin memnuniyet puanları; boşanmış, eşi ölmüş yada hiç evlenmemiş kişilere göre daha yüksektir.

Mutluluk düzeyi orta olan özel sektörde çalışanların (Dügüm 5) memnuniyet puanları sadece cinsiyet değişkeni tarafından etkilenmektedir. (Düz. $P=0.000$). Bu sonuca göre mutluluk düzeyi orta olan özel sektörde çalışan erkeklerin memnuniyet puanları; kadınlara göre daha yüksektir.

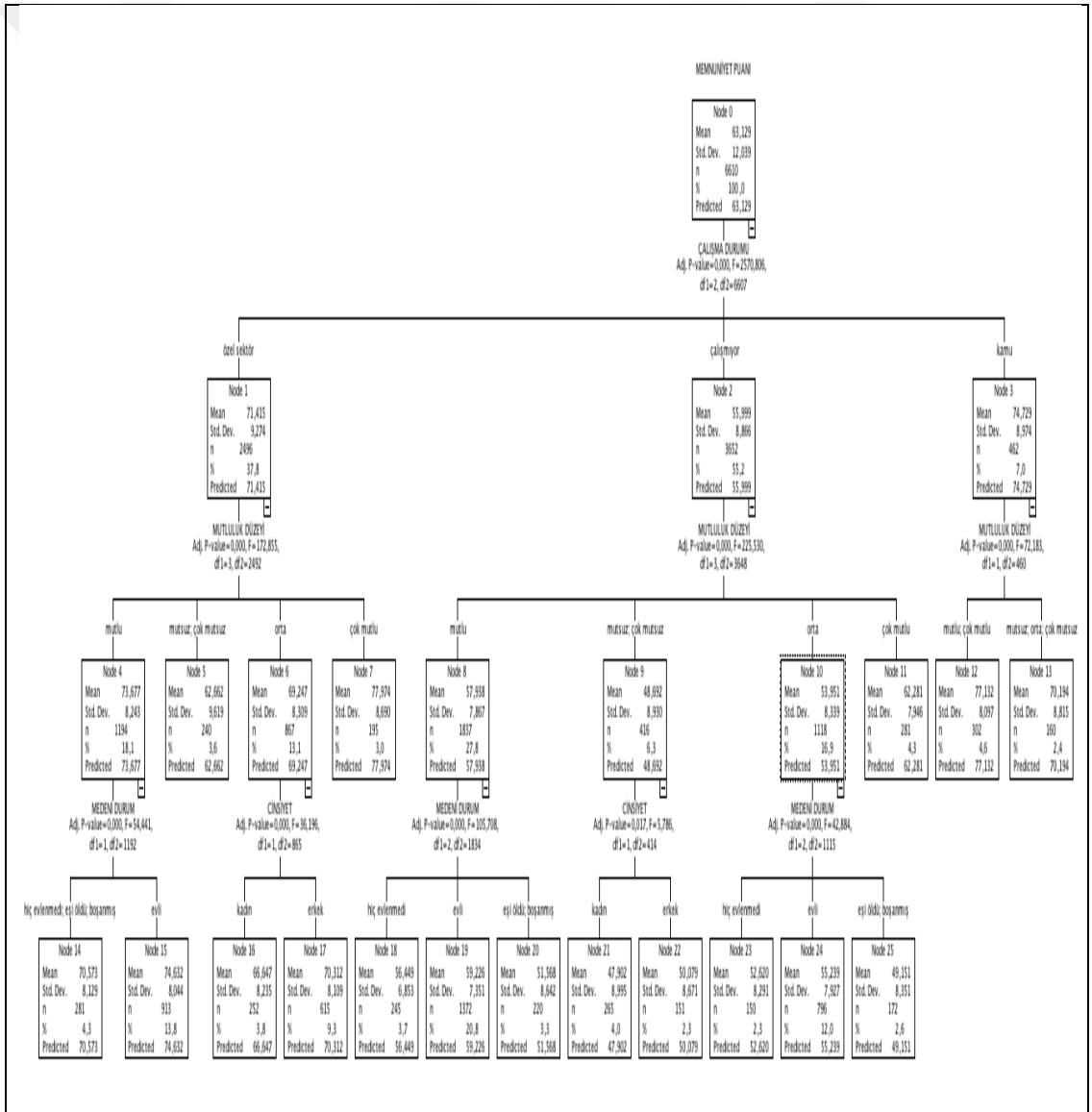
Dügüm 2'de çalışmayanların memnuniyet puanları, mutluluk düzeyi değişkeni tarafından etkilenmiştir (Düz $P=0.000$). Dügüm 2 (çalışmayan kişiler) mutluluk düzeyi bakımından dört yeni alt gruba (Dügüm 7,8,9 ve 10) ayrılmıştır. Dügüm 7; mutluluk düzeyi, mutlu olan çalışmayanların oluşturduğu alt grubu, Dügüm 8; mutluluk düzeyi orta olan çalışmayanların oluşturduğu alt grubu, Dügüm 9; mutluluk düzeyi mutsuz ve çok mutsuz olan çalışmayanların oluşturduğu alt grubu ve Dügüm 10; mutluluk düzeyi çok mutlu olan çalışmayanların oluşturduğu alt grubu ifade etmektedir. Dügüm 7, Dügüm 8, Dügüm 9 ve Dügüm 10 için memnuniyet puan ortalamaları sırasıyla, 57,98 ($S=7,9$) , 53,8 ($S=8,3$), 48,63 ($S=8,9$) ve 61.7 ($S=7,98$) olarak tahmin edilmiştir.

Mutluluk düzeyi, mutlu olan çalışmayanların (Dügüm 7) memnuniyet puanları sadece medeni durum tarafından etkilenmektedir. (Düz. $P=0.000$). Bu sonuca göre mutluluk düzeyi, mutlu olan ve çalışmayan kişilerde evlilerin memnuniyet puanları; boşanmış, eşi ölmüş yada hiç evlenmemiş kişilere göre daha yüksektir.

Mutluluk düzeyi, orta olan çalışmayanların (Dügüm 8) memnuniyet puanları sadece medeni durum tarafından etkilenmektedir. (Düz. $P=0.000$). Bu sonuca göre mutluluk düzeyi, orta olan ve çalışmayan kişilerde evlilerin memnuniyet puanları; boşanmış, eşi ölmüş yada hiç evlenmemiş kişilere göre daha yüksektir.

CHAID algoritması ile oluşturulan ikinci regresyon ağacına bakıldığı zaman kişilerin memnuniyet puanlarını etkileyen en önemli değişkenin çalışma durumu olduğu görülmüştür. Kamuda çalışanların memnuniyet puanları özel sektörde çalışanlardan, özel sektörde çalışanların çalışmayanlara göre memnuniyet puanları daha yüksektir. Memnuniyet puanlarını çalışma durumundan sonra mutluluk düzeyi ve medeni durumları ve cinsiyet etkilemektedir. Çalışmayan ve mutlu olan evli kişilerin diğer medeni durumdakilere göre memnuniyet puanının daha yüksek olduğu belirlenmiştir. Özel sektörde çalışan veya çalışmayan ve mutlu kişilerde evlilerin memnuniyet puanları; boşanmış, eşi ölmüş yada hiç evlenmemiş kişilere göre daha yüksektir. Özel sektörde çalışan ve mutluluk düzeyi orta olan erkeklerin memnuniyet puanları; kadınlara göre daha yüksektir.

CHAID algoritması ile üçüncü model oluşturulurken veri setinin %70'i modeli oluşturmak, %30'u da modeli test etmek için kullanılmıştır. Bunun sonucunda oluşturulan ağaç şekli 24'te gösterildiği şekildedir.



Şekil 24. CHAID Algoritması İle Oluşturulan 3. Regresyon Ağacı

CHAID modeli ile oluşturulan üçüncü regresyon ağacı diyagramı incelendiğinde, memnuniyet puanları üzerinde birinci derecede etkili bağımsız değişkenin çalışılan yer (Düz. $P=0.000$), ikinci derecede etkili bağımsız değişken mutluluk düzeyi (Düz. $P=0.000$), üçüncü derecede etkili bağımsız değişkenlerin Medeni Durum (Düz. $P=0.000$) ve cinsiyet (Düz. $P=0.000$, Düz. $P=0.017$) olduğu belirlenmiştir.

Memnuniyet puanları (Düğüm 0), birinci derecede etkili olan çalışılan yer değişkeni tarafından 3 alt gruba ayrılmıştır. Düğüm 1; özel sektörde çalışanların oluşturduğu alt grubu, Düğüm 2; çalışmayan kişilerin oluşturduğu alt grubu, Düğüm 3; kamuda çalışanların oluşturduğu alt grubu temsil etmektedir. Düğüm 1 (özel sektörde çalışanların) tmenuniyet puanı ortalaması 71,4 ($S=9,3$), Düğüm 2 (çalışmayanların) memnuniyet puan ortalaması 55,99 ($S=8,86$), Düğüm 3 (kamuda çalışanların) memnuniyet puan ortalaması 74,7 ($S=8,9$) olarak bulunmuştur .

Oluşturulan alt gruplardan yeterince homojen yapıya sahip olan Düğüm 3'e terminal düğüm adı verilir. Ayrıca memnuniyet puan ortalamalarına bakıldığında memnuniyet puanları sırasıyla kamuda çalışanlar, özel sektörde çalışanlar ve çalışmayanlar olarak belirlenmiştir.

Düğüm 1'de özel sektörde çalışanların memnuniyet puanları, mutluluk düzeyi değişkeni tarafından etkilenmiştir (Düz $P=0.000$). Düğüm 1 (özel sektörde çalışan kişiler) mutluluk düzeyi bakımından dört alt gruba (Düğüm 4, 5, 6 ve 7) ayrılmıştır. Düğüm 4; mutluluk düzeyi mutlu olan özel sektörde çalışanların oluşturduğu alt grubu, Düğüm 5; mutluluk düzeyi mutsuz ve çok mutsuz olan özel sektörde çalışan kişilerin oluşturduğu alt grubu, Düğüm 6; mutluluk düzeyi orta olan özel sektörde çalışan kişileri ve Düğüm 7; mutluluk düzeyi çok mutlu olan özel sektörde çalışan kişileri ifade etmektedir. Düğüm 4, 5, 6 ve 7 için memnuniyet puan ortalamaları sırasıyla, 73,7 ($S=8,2$), 62,7 ($S=9,6$), 69,2 ($S=8,3$) ve 77,97 ($S=8,7$) olarak tahmin edilmiştir.

Mutluluk düzeyi mutlu olan özel sektörde çalışanların (Düğüm 4) memnuniyet puanları sadece medeni durum tarafından etkilenmektedir (Düz. $P= 0.000$). Bu sonuca göre mutluluk düzeyi mutlu olan özel sektörde çalışan kişilerde evlilerin memnuniyet puanları, boşanmış, eşi ölmüş yada hiç evlenmemiş kişilere göre daha yüksektir. Mutluluk düzeyi orta olan özel sektörde çalışanların (Düğüm 6) memnuniyet puanları sadece cinsiyet değişkeni tarafından etkilenmektedir (Düz. $P= 0.000$). Bu sonuca göre mutluluk düzeyi mutlu olan, özel sektörde çalışan erkeklerin memnuniyet puanları, kadınlara göre daha yüksektir.

Düğüm 2'de çalışmayanların memnuniyet puanları analize dahil edilen bağımsız değişkenlerden sadece mutluluk düzeyi tarafından etkilenmektedir (Düz. $P=0.000$). Düğüm 2; mutluluk düzeyi bakımından dört yeni alt gruba (Düğüm 8, 9, 10 ve 11) ayrılmıştır. Düğüm 8; çalışmayan ve mutluluk düzeyi mutlu olan kişilerin oluşturduğu alt grubu, Düğüm 9; çalışmayan ve mutluluk düzeyleri mutsuz, çok mutsuz kişilerin oluşturduğu alt grubu, Düğüm 10; çalışmayan ve mutluluk düzeyleri orta olan kişilerin oluşturduğu alt grubu, Düğüm 11; çalışmayan ve mutluluk düzeyleri çok

mutlu olan kişilerin oluşturduğu alt grubu ifade etmektedir. Düğüm 8, 9, 10 ve 11 için memnuniyet puan ortalamaları sırasıyla, 57,9 (S=7,9), 48,7 (S=8,9), 53,9 (S=8,3) ve 62,3 (S=7,9) olarak hesaplanmıştır.

Mutluluk düzeyi mutlu olan çalışmayanların (Düğüm 8) memnuniyet puanları sadece medeni durum tarafından etkilenmektedir (Düz. P= 0.000). Düğüm 8; medeni durum değişkeni tarafında üç alt düğüme (Düğüm 18, 19 ve 20) ayrılmıştır. Düğüm 18; çalışmayan, mutlu, hiç evlenmemiş olan kişilerin oluşturduğu alt grubu, Düğüm 19; çalışmayan mutlu evli kişilerin oluşturduğu alt grubu, Düğüm 20; çalışmayan, mutlu, eşi ölmüş yada boşanmış kişilerin oluşturduğu alt grubu ifade etmektedir. Düğüm 18, 19 ve 20 için memnuniyet puan ortalamaları sırasıyla, 56,45 (S=6,8), 59,2 (S=7,3) ve 52,6 (S=8,6) olarak hesaplanmıştır. Bu sonuca göre mutluluk düzeyi mutlu olan çalışmayan kişilerde evlilerin memnuniyet puanları, hiç evlenmemiş kişilerden; hiç evlenmemiş kişilerin memnuniyet puanları boşanmış, eşi ölmüş kişilerden daha yüksektir.

Mutluluk düzeyi mutsuz, çok mutsuz olan çalışmayanların (Düğüm 9) memnuniyet puanları sadece cinsiyet değişkeni tarafından etkilenmektedir (Düz. P= 0.017). Bu sonuca göre mutluluk düzeyi mutlu ve çok mutlu olan, çalışmayan erkeklerin memnuniyet puanları, kadınlara göre daha yüksektir.

Mutluluk düzeyi orta olan çalışmayanların (Düğüm 10) memnuniyet puanları sadece medeni durum tarafından etkilenmektedir (Düz. P= 0.000). Düğüm 10; medeni durum değişkeni tarafında üç alt düğüme (Düğüm 23, 24 ve 25) ayrılmıştır. Düğüm 23; çalışmayan, mutluluk düzeyi orta olan, hiç evlenmemiş kişilerin oluşturduğu alt grubu, Düğüm 24; çalışmayan, orta düzey mutlu evli kişilerin oluşturduğu alt grubu, Düğüm 25; çalışmayan, orta düzey mutlu, eşi ölmüş yada boşanmış kişilerin oluşturduğu alt grubu ifade etmektedir. Düğüm 23, 24 ve 25 için memnuniyet puan ortalamaları sırasıyla, 52,6 (S=8,3), 55,2 (S=7,9) ve 49,1 (S=8,3) olarak hesaplanmıştır. Bu sonuca göre mutluluk düzeyi mutlu olan çalışmayan kişilerde evlilerin memnuniyet puanları, hiç evlenmemiş kişilerden; hiç evlenmemiş kişilerin memnuniyet puanları boşanmış, eşi ölmüş kişilerden daha yüksektir.

CHAID algoritması ile oluşturulan regresyon ağacı modellerinde yaşam memnuniyet puanlarına bakıldığı zaman memnuniyet puanlarını etkileyen en önemli değişkenin kişilerin çalışma durumu olduğu görülmüştür. Ağaçların hepsinde çıkan sonuç kamuda çalışan insanların yaşam memnuniyet puanlarının daha yüksek olduğudur. Model veri oranı arttıkça kişilerin memnuniyet puanlarını etkileyen değişkenlerin mutluluk düzeyleri, medeni durumları ve cinsiyetleri olduğu görülmüştür. Genel olarak 3 ağaç incelendiğinde, kişiyi kimin mutlu ettiği ve neyin mutlu ettiği kişilerin memnuniyet puanını etkilememektedir. Ayrıca, memnuniyet puanını 1. dereceden ve 2. dereceden etkileyen değişkenler farklılık göstermemektedir. Ancak ağaçlar arasındaki en önemli farklılık model verisi oranı arttıkça bağımlı değişkeni etkileyen bağımsız değişken sayısının ve düğüm sayılarının artmasıdır. Bu sonuçta, ağacın daha karmaşık olmasını sağlasada model hakkında detaylı bilgi vermektedir. Dolayısıyla ağaç belirlenirken model verisinin en yüksek olduğu ağacın tercih edilmesi önerilmektedir.

4 SONUÇ

CART ve CHAID algoritmaları hem kategorik hemde sürekli değişkenler kullanılarak sınıflama ve regresyon ağacı problemlerinin çözümünde kullanılmaktadır. Bağımlı değişkenin sürekli olması durumunda kestirim amaçlı olan bu yöntemler, bağımlı değişkenin kategorik olması durumunda sınıflandırma amacına sahiptir.

Bu çalışmada; TÜİK tarafından gerçekleştirilen yaşam memnuniyeti anketi verileri doğrultusunda, kişilerin mutluluk düzeylerini etkileyen faktörler sınıflama ve regresyon ağaçları ile belirlenmeye çalışılmıştır. Sınıflama ve regresyon ağaçlarında CART ve CHAID algoritmaları kullanılmıştır. Ayrıca sınıflama ve regresyon ağaçları, model verisinin %30, %50 ve %70'lik kısmı dahil edilerek oluşturulmuş ve oluşturulan ağaçlar karşılaştırmalı olarak yorumlanmıştır.

Çalışmanın sonucunda oluşturulan sınıflama ağaçlarında;

Kişilerin mutluluk düzeylerini etkileyen en önemli faktörün gelecekte umutlu olma olduğu sonucuna ulaşılmıştır. Buna göre; umut seviyesi düştükçe mutluluk oranları da azalmakta olduğu sonucuna ulaşılmıştır. Geleceğinden umutlu olan insanlarda mutlu olanların oranı, geleceğinden umutlu olmayanlarda orta düzey veya mutsuzların oranı daha yüksektir.

Aynı zamanda geleceğinden umutlu olup sosyal hayatından da memnun olan insanların sosyal hayatından memnun olmayan insanlara göre çok daha yüksek oranda mutlu olduğu sonucuna ulaşılmıştır.

Kişiler için mutluluk düzeylerini umuttan sonra etkileyen değişkenlerin evlilik ve gelir olduğu görülmüştür. Evliliğinden yada gelirinden memnun olan kişilerin memnun olmayan kişilere göre çok daha mutlu olduğu sonucuna ulaşılmıştır.

Mutluluk seviyeleri; geleceğinden umutlu olmayıp gelirinden de memnun olmayan kadınlarda orta seviyede iken, hem geleceğinden umutlu olmayıp hemde gelirinden memnun olmayan erkeklerde mutsuzluk seviyesi yüksektir. Bu durum erkeklerin gelire kadınlardan daha fazla önem verdiğini göstermektedir.

CART ve CHAID algoritması ile oluşturulmuş sınıflama ağaçları incelendiğinde mutluluğu etkileyen en önemli değişkenin değişmediği görülmüştür. Ayrıca her iki algoritmada da mutluluğu etkileyen bağımsız değişkenler (evlilikten ve gelirden memnun olma) aynı çıkmış sadece model veri oranına göre yeni bağımsız değişkenlerin arttığı (sosyal hayattan ve kendine ayırdığı zamandan memnuniyet) görülmüştür.

Model verisi arttıkça mutluluğu etkileyen değişken sayısı ve düğüm sayısı da artmaktadır ve ağaç daha karmaşık bir yapıya bürünmektedir. Dolayısıyla, kategoriler yani sınıflar hakkında daha ayrıntılı bilgiye ulaşılmaktadır. Sınıflar arasındaki etkileşimlerde görülebilmektedir. Bu durumda da araştırmacı analiz ettiği veriye ait daha detaylı bilgiye sahip olacağından model verisinin yüksek tutulması önerilmektedir.

Çalışmanın sonucunda oluşturulan regresyon ağaçlarında; CART algoritması ile oluşturulan regresyon ağaçları incelendiğinde;

Memnuniyet puanlarını etkileyen en önemli değişkenin kişilerin çalışma durumları olduğu belirlenmiştir.

Oluşturulan modeller sonucunda özel sektörde ya da kamuda çalışan kişilerin mutluluk düzeyleri ne olursa olsun genel olarak çalışmayanlara göre memnuniyet puanlarının daha yüksek, yani hayatlarından daha memnun olduğu sonucuna ulaşılmıştır.

Mutluluk düzeylerine bakıldığında çalışan kişilerde de çalışmayan kişilerde de mutlu ve çok mutlu olanların memnuniyet puanları, mutluluk düzeyi orta, mutsuz, çok mutsuz olanlara göre daha yüksektir.

Model veri oranı arttıkça memnuniyet puanlarını; mutluluk düzeyleri, medeni durum ve cinsiyet değişkenlerinin de etkilediği sonucuna ulaşılmıştır.

Özel yada kamuda çalışan kendini mutlu hisseden evli kişilerin memnuniyet puanları diğer medeni durumdaki kişilere göre oldukça yüksektir.

Evli ve hiç evlenmemiş mutlu erkeklerin memnuniyet puanları ise evli ve hiç evlenmemiş mutlu kadınların memnuniyet puanlarından biraz daha yüksek olduğu görülmüştür.

Kişiyi kimin mutlu ettiğinin ve neyin mutlu ettiğinin kişilerin memnuniyet puanını etkilemediği görülmüştür.

Ayrıca model veri oranı artsa da, memnuniyet puanını 1. dereceden ve 2. dereceden etkileyen değişkenler farklılık göstermediği sonucuna ulaşılmıştır. Ancak ağaçlar arasındaki en önemli farklılık model verisi oranı arttıkça bağımlı değişkeni etkileyen bağımsız değişken sayısının ve düğüm sayılarının artmasıdır. Dolayısıyla, kategoriler yani sınıflar hakkında daha ayrıntılı bilgiye ulaşılmaktadır. Sınıflar arasındaki etkileşimlerde görülebilmektedir. Bu durumda, ağaç belirlenirken model verisinin en yüksek olduğu ağacın tercih edilmesi önerilmektedir.

CHAID algoritması ile oluşturulan regresyon ağaçları incelendiğinde;

Memnuniyet puanlarını etkileyen en önemli değişkenin kişilerin çalışma durumu olduğu görülmüştür. Kamuda çalışan kişilerin memnuniyet puanlarının özel sektörde çalışanlardan, özel sektörde çalışan kişilerin memnuniyet puanlarının ise çalışmayanlara göre daha yüksektir.

En yüksek memnuniyet puanına sahip kişiler kamuda çalışanlardır.

Model veri oranı arttıkça kişilerin memnuniyet puanlarını etkileyen değişkenlerin mutluluk düzeyleri, medeni durumları ve cinsiyetleri olduğu da görülmüştür.

Çalışmayan ve mutlu olan evli kişilerin diğer medeni durumdakilere göre memnuniyet puanının daha yüksek olduğu belirlenmiştir.

Özel sektörde çalışan ve mutlu olan erkeklerin ise kadınlara göre memnuniyet puanları daha yüksek olduğu görülmüştür.

Genel olarak 3 ağaç incelendiğinde, kişiyi kimin mutlu ettiği ve neyin mutlu ettiği kişilerin memnuniyet puanını etkilememektedir. Ayrıca, regresyon ağaçlarında model veri oranı değişmede bağımlı değişkeni etkileyen değişkenlerin ve önem sırasının değişmediği görülmüştür. Ancak ağaçlar arasındaki en önemli farklılık model verisi oranı arttıkça bağımlı değişkeni etkileyen bağımsız değişken sayısının ve düğüm sayılarının artmasıdır. Bu sonuçta, ağacın daha karmaşık olmasını sağlasada model hakkında detaylı bilgi vermektedir. Dolayısıyla ağaç belirlenirken model verisinin en yüksek olduğu ağacın tercih edilmesi önerilmektedir.

Araştırma sonuçlarında oluşan ağaçlar ve hangi ağacın kullanılması gerektiği araştırmanın konusuna bağlı olarak araştırmacının tercihine bırakılmalıdır. Daha detaylı sonuçlar için model verisinin yüksek tutulması daha kısa sonuçlar için ise model veri oranının düşük tutulması önerilmektedir. Ağaç seçimi hatalı sınıflama oranı veya maliyet karmaşıklık hesabı kullanılarak da belirlenebilmektedir.

Sonuç olarak, CART ve CHAID algoritmalarının ağacı oluşturma ve geliştirme yöntemleri birbirinden farklı olduğundan iki ağaç arasında farklılıklar görülebilmektedir. CHAID algoritmasında düğümler ikiden fazla sınıflara ayrılabilirken CART algoritmasında sınıflar sadece ikili olarak ayrılmaktadır. Dolayısıyla sınıflandırma ağacında CHAID algoritması daha kapsamlı sonuçlara ulaşmayı sağlarken CART algoritmasında daha genel sonuçlar elde edilir. Araştırmacı hangi algoritmanın daha uygun olduğuna yorumlama aşamasında kendi karar verebilmektedir.

Kaynakça

- (2007, 09 09). 03 21, 2017 tarihinde İstatistik / Kütüphane: www.stat.gen.tr adresinden alındı
- Akçapınar Sezer , E., Bozkır, A. S., Yağız, S., & Gökçeoğlu, S. (2010). Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinasının İlerleme Hızı Üzerinde Uygulama. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*. Kayseri.
- Akkartal, E., & Mendes, M. (2009). Regression Tree Analysis For Predicting Slaughter Weight İn Broilers. *Italian Journal of Animal Science*, 615-624.
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 1-22.
- Albar Özbucak, B. (2014). Tüketicilerin Özel Markalı Ürünleri Satın Almasına Tutundurma Stratejilerinin Etkisi: Giresun İlinde Bir Uygulama. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 99-115.
- Altaş, D., & Gülpınar, V. (2012). Karar Ağaçları ve Yapay Sinir Ağlarının Sınıflandırma Performanslarının Karşılaştırılması: Avrupa Birliği Örneği. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 14(1).
- Atılğan, E. (2011). Karayollarında Meydana Gelen Trafik Kazalarının Karar Ağaçları ve Birliktelik Analizi İle İncelenmesi. *Yüksek Lisans Tezi*. Hacettepe Üniversitesi İstatistik Anabilim Dalı.
- Babaoğlu, H. (2008, 05 09). *VATAN*. 04 25, 2017 tarihinde m.gazetevatan.com adresinden alındı
- Baran, B., & Ata, F. (2014). Üniversite Öğrencilerinin Bilgi Okuryazarlık Öz Yeterlilik Algısının Karar Ağacı Analizi İle İncelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 137-160.
- Baysal, A. (2011). CHAID ve Regresyon Ağacı Analiz Yöntemlerinin Hemşirelerde Stres Düzeyini Etkileyen Faktörler Üzerine Uygulaması. *Atatürk Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi*. Erzurum.
- Beşel, F. (2015). 2013 Yılı Yaşam Memnuniyeti Araştırması Sonuçlarının İl Bazlı Ekonomik, Sosyal ve Siyasi Analizi. *Sakarya Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*.
- Bozan, F. (2010). 3 12, 2017 tarihinde www.farukbozan.com/2010/01/cartclassification-and-regression-tree/ adresinden alındı
- Bozkuş, S., Çevik, E. İ., & Üçdoğruk, Ş. (2006). Bireysel Refah ve Mutluluk Düzeyine Etki Eden Faktörlerin Sıralı Logit İle Modellenmesi: Türkiye Örneği. İzmir: 15. İstatistik Araştırma Sempozyumu.
- Breiman, L., Freidman, J. H., Olshen, A. R., & Stone, J. C. (1993). Classification and Regression Trees. *New York: Chapman & Hall*.
- Breiman, L., Freidman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. *CRC Press*.
- Breiman, L., Freidman, J., Olshen, R., & Stone, C. J. (1993). Classification and Regression Trees. *New York: Chapman & Hall*.
- Bülbül, Ş., & Giray, S. (2011). Sosyodemografik Özellikler ile Mutluluk Algısı Arasındaki İlişki Yapısının Analizi. *Ege Akademik Bakış*, 113-123.
- Chow, H. (2005). Life Satisfaction among University Students in a Canadian Prairie City: A Multivariate Analysis. *Social Indicators Research*, 70, s. 139-150.

- Çölkesen, R. (2014). *Veri Yapıları ve Algoritmalar*. İstanbul: Papatya Yayıncılık Eğitim.
- De'ath, G., & Fabricius, K. (2008). Classification and regression trees:a powerful yet simple technique for eqological data analysis. *Eceology*, 3178-3192.
- Doğan, İ. (2003). Holştayn Irkı İneklerde Süt Verimine Etki Eden Faktörlerin Chaid Analizi İle İncelenmesi. *Ankara Üniversitesi Veterinerlik Fakültesi Dergisi*, 65-70.
- Ersöz, T., Özseven, T., & Ersöz, F. (2017). Tüketicilerin Cep Telefonu Tercihlerinin Karar Ağacı İle Modellenmesi. *Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Elektronik Dergisi*.
- Gilman, R., Huebner, E. S., & Laughlin, J. E. (2000). A First Study of the Multidimensional Students Life Satisfaction Scale with Adolescents. *Social İndicators Research*, 52, s. 135-160.
- Giray, S. (2011). Doğrusal Olmayan Kanonik Korelasyon Analizi ve Yaşam Memnuniyeti Üzerine Bir Uygulama. *Doktora Tezi*. İstanbul: Marmara Üniversitesi.
- Güler, B. K., & Emeç, H. (2006). Yaşam Memnuniyeti ve Akademik Başarıda İyimserlik Etkisi. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi*, 129-149.
- Günüç, S. (2013). İnternet Bağımlılığını Yotdayan Bazı Değişkenlerin Cart ve Chaid Analizleri ile İncelenmesi. *Türk Psikoloji Dergisi*, 88-101.
- IBM SPSS Modeler 15 Algorithms Guide*. (tarih yok). 2 25, 2017 tarihinde <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/en/AlgorithmsGuide.pdf> adresinden alındı
- Kahyaoglu, O. (2008). Yaşam Memnuniyeti ve Yaşam Memnuniyetini Etkileyen Değişkenler İle Ekonometrik Uygulama: Türkiye Örneği. *Yüksek Lisans Tezi*. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.
- Kayri, M., & Boysan, M. (2007). Araştırmalarda Chaid Analizinin Kullanımı ve Baş Etme Stratejileri İle İlgili Bir Uygulama. *Hacettepe Üniversitesi Eğitim Bilimleri Dergisi*, 133-149.
- Kayri, M., & Boysan, M. (2008). Assesment of relation between cognitive vulnerability and depression's level by using classification and regression tree analysis. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 168-177.
- Kayri, M., & Boysan, M. (2008). Bilişsel Yatkınlık İle Depresyon Düzeyleri İlişkisinin Sınıflandırma ve Regresyon Ağacı İle İncelenmesi. *Hacettepe Üniversitesi Eğitim Bilimleri Dergisi*, 168-177.
- Kıran, Z. B. (2010, 05). Lojistik Regresyon ve CART Analizi Teknikleriyle Sosyal Güvenlik Kurumu İlaç Provizyon Sistemi Verileri Üzerinde Bir Uygulama. *Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.
- Koç, Y. (2016). Hsyvsnclıkta Elde Edilen Farklı Verilerle Regresyon Ağacı Metodunun Uygulanması. *Yüksek Lisans Tezi*. Iğdır: Zootekni Anabilim Dalı.
- Köksal, B. (2011). Regresyon Analizinde ROC Eğrisi Kestirimi İle Model Seçimi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı.
- Korkmaz, M., Yücel, A. S., Gürkan, A., & Germir, H. N. (2015). Yaşam Memnuniyeti Üzerinde Etkili Olan Sosyodemografik Bileşenler Üzerinde Bir Analiz. *Uluslararası Hakemli Psikiyatri ve Psikoloji Araştırma Dergisi*.

- Koyuncugil, A. S., & Özgülbaş, N. (2008). İMKB'de İşlem Gören Kobilerin Güçlü ve Zayıf Yönleri:Chaid Karar Ağacı Uygulaması . *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi*, 1-21.
- Küçüköğlü, O. (2010). Veri Madenciliği Yöntemlerinin Hayvancılıkta Kullanımı. Çanakkale: Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsü.
- Lewis, R. j. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. *Presented at the 2000 annual meeting of the society for Academic Emergency Medicine in San Francisco*, s. 1-14.
- Mertens, M., Nestler, I., & Huwe, B. (2002). GIS-based regionalization of soil profiles with classification and regression trees. *journal of plant nutrition and soil science*, s. 39-43.
- Morgan, N. J., & Sonnquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 415-434.
- Nisbet, R., Elder , J., & Miner, G. (2009). Handbook of statistical analysis and data mining applications. *Elsevier*.
- Oğuzlar, A. (2004). CART Analizi ile Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi. *İktisadi ve İdari Bilimler Dergisi*.
- Oğuzlar, A. (2004). CART Analizi ile Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi. *İktisadi ve İdari Bilimler Dergisi*, 79-90.
- Oruçoğlu, O. (2011). Holstein Irkı İneklerin 305 Günlük Süt Verimini Etkileyen Çevre Faktörlerinin Regresyon Ağacı İle Belirlenmesi. *Yüksek Lisans Tezi*. Isparta: Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü.
- Özkan, K. (2012). Sınıflandırma ve Regresyon Ağacı Tekniği (SRAT) ile Ekolojik Verinin Modellenmesi. *Süleyman Demirel Üniversitesi Orman Fakültesi Dergisi*, 31-52.
- Özkan, K. (2012). Sınıflandırma ve Regresyon Ağacı Tekniği(SRAT) ile ekolojik verinin modellenmesi. *SDÜ Orman Fakültesi Dergisi*, 1-4.
- Özkan, Y. (2016). *Veri Madenciliği Yöntemleri*. Papatya Bilim.
- Özkan, Y. (2016). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık.
- Rokach, L., & Maimon, O. (tarih yok). *Decision Trees*. 2 23, 2017 tarihinde <http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf> adresinden alındı
- Saitoğlu, Y. S. (2015). Sınıflama ve Regresyon Ağaçları. *Doktora Tezi*. İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.
- Seidman, C. (2001). *Data Mining with Microsoft SQL Server 2000*. Washington: Microsoft Press.
- Şeker, M. (2009). Mutluluk Ekonomisi. *Sosyoloji Konferansları Dergisi*.
- Sezer, E. A., Bozkır, A. S., Yağız, s., & Gökçeoğlu, C. (2010). Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu.
- Sezgin, Ö. (2006). *Statistical Method in Credit Banking*.
- Silahtaroglu, G. (2013). *Veri Madenciliği (Kavram ve Algoritmaları)* (Papatya Yayıncılık b.). İstanbul.
- Soner, V. (2015, 09). *Horizon 2020&TUBITAK/İstatistik/Yöneylem Araştırması*. 02 11, 2017 tarihinde <http://volkansoner.wordpress.com/pace/11> adresinden alındı
- Sutton, C. D. (2005). Classification and Regression Trees, Bagging, and Boosting. C. R. Rao, E. J. Wegman, & J. L. Solka içinde, *Handbook of statistics 24* (s. 303-331). ELSEVIER.

- Temel, G. O. (2004). Sınıflama ve Regresyon Ağaçları. *Yüksek Lisans Tezi*. mersin: mersin üniversitesi sağlık bilimleri enstitüsü biyoistatistik anabilim dalı.
- Timofeev, R. (2004, December 20). Classification and Regression Trees(CART) Theory and Application. *A Master Thesis Presented*. Berlin: CASE-Center of Applied Statistics and Economics Humboldt University.
- Türkiye İstatistik Kurumu*. (2017). 2017 tarihinde Türkiye İstatistik Kurumu: www.tuik.gov.tr adresinden alındı
- Yohannes, Y., & Hoddinot, J. (1999). Classification and Regression Trees: An Introduction. *International Food Policy Research Institute*.
- Yohannes, Y., & Webb, P. (1999). *Classification and Regression Trees, Cart*. Microcomputers In Policy Research3.



ÖZGEÇMİŞ

Adı Soyadı : Yasemin Bahar YÜCEL
Doğum Yeri ve Yılı : İSTANBUL, 28/05/1992
Medeni Hali : (Bekar)
Yabancı Dili : İngilizce
E-posta : yucelyaseminbahar@gmail.com

Eğitim Durumu

Lise : Cibali Lisesi, 2010
Lisans : İstanbul Ticaret Üniversitesi, /// Fen-Edebiyat Fakültesi,
/// İstatistik Bölümü
Yüksek Lisans : İstanbul Ticaret Üniversitesi,
Fen Bilimleri Enstitüsü, /// İstatistik Anabilim Dalı

Mesleki Deneyim

DENİZBANK,
Hazine Departmanı Stajyeri 2013
BAREM RESEARCH,
Kantitatif Departmanı Stajyeri 2014

Yayınları

Kalkan, B.S., Yücel, Y.B., 2017. Determination of Factors Affecting Happiness Level by Classification Tree Technique. European Journal of Business and Social and Sciences, (6)