T.C. İSTANBUL TİCARET ÜNİVERSİTESİ

DIŞ TİCARET ENSTİTÜSÜ
WORKING PAPER SERIES

Tartışma Metinleri

WPS NO/ 165/2018-05

KALP HASTALIKLARINA ETKİ EDEN BAZI FAKTÖRLERİN CART VE CHAID TEKNİĞİ İLE BELİRLENMESİ

Onur KÖSE*

*onur.kose@ito.org.tr İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü İstatistik Tezli Yüksek Lisans Programı Öğrencisi

Özet

Karar ağaçlarının algoritmalarından olan CHAID ve CART teknikleri; kullanımının ekonomik olması ve hızlı sonuç vermesi nedeniyle, veri işlemede sıkça kullanılan tekniklerdendir. Bu çalışmada; University of California bünyesinde veri setlerini barındıran bir platformdan alınan kalp hastalığına etki eden 38 adet faktör kullanılmıştır. Bu faktörlerin değerlendirilmesi; Sınıflama ve Regresyon Ağaçı (CART) ve Otomatik Ki-Kare Etkileşim Belirleme (CHAID) algoritmaları kullanılarak oluşturulmuştur ve çıkan sonuçlar birbirleriyle karşılaşırılarak yorumlanmıştr.

Anahtar kelimeler: Classification and Regression Tree (CART) ve Chi-Squared Automatic Interaction Detector (CHAID), Kalp hastalığı.

Abstract

CHAID and CART techniques which are the algorithms of desicion trees and are frequently used techniques in terms of getting quick result and being economic. In this study, 38 factors related to heart diseases are studied and data were taken from a platform which has data sets in University of California. These factors were evaluated by Classification and Regression Tree (CART) and Chi-Squared Automatic Interaction Detector (CHAID) algorithms and the results were interpreted by comparing to one another

Keywords: Classification and Regression Tree (CART) and Chi-Squared Automatic Interaction Detector (CHAID), heart disease.

Giriş

Karar ağaçları ilk olarak; 1973 yılında Bierman ve Friedman tarafından geliştirilmiş olup değişkenlerin parçalanarak bir ağaç oluşturulması prensibine dayanmaktadır (Ulusoy, 2013). Karar ağaçları, sınıflandırma amacıyla veri madenciliğinde en çok kullanılan tahmin edici bir tekniktir. Genellikle sınıflandırma, kümeleme, tahmin modellerinde ve sorunla ilgili araştırma alanını alt gruplara ayırmak için kullanılmaktadır (Quinlan, 1986).

Karar ağacı; kök düğüm, dallar ve yapraklardan oluşur. Bu tipik bir ağacı andıran yapıda; karar düğümleri, yapılacak testi belirtir. Buradaki amaç; ağacın veri kaybetmeden dallara ayrılmasıdır. Her düğümde test ve dallara ayrılma işlemleri ardışık olarak gerçekleşir. Bu ayrılma işlemi üst seviyedeki ayrımlara bağımlıdır. Ağaca ait her dal; sınıflama işlemini tamamlamaya yönelik hareket eder. Eğer bir dalın ucunda sınıflama işlemi gerçekleşemiyorsa o dalın ucunda bir karar düğümü oluşur ki buna yaprak düğüm denir. Bu yaprak düğüm, veri üzerinde belirlenmek istenen sınıflardan birini ifade eder (Özekes, 2002: 21).

Sınıflama ve Regresyon Ağaçları Tekniği (CART) (Classification and Regression Trees) sürekli veya kategorik bağımlı değişkenlerin sayısal karşılıklarını öngörebilmek ve çözümleyebilmek amacıyla oluşturulmuş, dağılımdan bağımsız istatistiksel yöntemlerdendir. CRAT, kategorik bağımlı değişkenlerde sınıflama ağacı şeklinde, sürekli bağımlı değişkenlerde ise regresyon ağacı şeklinde adlandırılmaktadır (Fu, 2004). CART modelleri, yinelenebilen tahmin ediciler evreninin eş tekrarlı iki alt sınıfa ayrıştırılması temeline dayanan karar ağaçları oluştururlar (Chipman & McCulloch, 2000). Karar noktalarına ulaşıncaya dek, iki alt sınıfa ayırma işlemi sürdürülür.

Otomatik Ki-Kare Etkileşim Belirleme (CHAID) (Chi-Squared Automatic Interaction Detector) ise; günümüzde birçok bakımdan sağladığı avantajlar sebebiyle kullanım alanı geniş bir analiz tekniğidir. CHAID; kullanıldığı bütün hiyerarşik seviyelerde ikiden daha fazla dallanma gösterir, bütün ölçü birimleiyle çalışır ve bütün bunların yanında verilerin normal dağılıma uyma zorunluluğu yoktur. Bağımlı değişkenin sürekli olması durumunda F, kategorik olması durumunda χ^2 testler kullanılmaktadır. Bununla birlikte sürekli bağımsız değişkenler analizinde otomatik olarak kategorik değişkenlere dönüştürülür. CHAID algoritmasının ileri düzey tanımlamalarında Pearson χ^2 veya likelihood-ratio testleri uygulanabilir (AKPINAR H., 2017).

Bu çalışmanın amacı; yaşları 40 ila 77 arasında değişen 248 hasta üzerinde kalp hastalığına etki eden 38 adet faktörün etkileri incelenmiştir. Çalışmada kalp hastalığına etki eden tansiyon, sigara kullanımı, diyabet ve ilaç kullanımı gibi faktörler CART ve CHAID analizleriyle karşılaştırmalı olarak değerlendirilmiştir.

Kalp; vücutta göğüs boşluğunda orta hatta bulunan tepesi aşağı- sola bakan ve dolaşım sistemine ait bir organdır. İnsan kalbi 4 temel boşluktan oluşur. Bunlar sırasıyla atrium dextrum, ventriculus dexter, atrium sinistrum, ventriculus sinister 'dir (Cumhur, 2014). Kalbin çalışma mekanizmasında veya yapısında meydana gelen herhangi bir aksaklık ya da bozukluk sonucu çeşitli kalp hastalıkları meydana gelebilir. Ancak günümüzde; kalp hastalığı en çok kalbi besleyen koroner arterlerin ya da kalbe ait diğer damarların tıkanması sonucu oluşmaktadır. Bu damarların tıkanmasında rol oynayan faktörler; kişinin cinsiyeti, yaşı, sigara kullanımı, tansiyon hastalığı, şeker hastalığı, egersiz yapma durumu olabilir (Yusuf ve ark, 2001).

1. METODOLOJİ

1.1 Karar Ağaçlari

Karar ağacının yapısı bir şema yapısındadır. Bu şemada her değişken bir düğüm tarafından ifade edilir. Ağaç yapısı kısaca kök, dallar ve yapraklardan oluşur. En üst yapı kök, en son yapı yaprak ve bunların arasında kalan yapılar ise dal olarak nitelendirilir. Karar ağaçlarının oluşturulmasında en önemli kısım; hangi değişkenin ilk düğüm (kök düğüm) olacağının belirlenmesidir. İlk düğüm; çeşitli kriterler kullanılarak belirlenir (Atılgan ES., 2011, 21).

Karar ağaçları çoğu kez karmaşık bir görünümde olabilir. Ağaç oluşum sürecinde atılan önemli adımlardan biri de budama işlemidir. Budama yöntemiyle ağaçta bulunan ancak sonucu etkilemeyen ve sınıflamaya herhangi bir katkısı olmayan dalların çıkarılması sağlanır. Böylelikle ağaçta gereksiz detayların bulunması engellenir. Buradaki asıl amaç; ağaçta birçok düğüm ve dal oluşursa, ağacın alt dalları ve yapraklarına ulaşan veri sayısı da azalacağından, dolaylı olarak ağacın hassasiyetini azaltacaktır. Böylece bu durum engellenmiş olur (Silahtaroğlu G., 2013).

CART ALGORİTMASI

CART (Classification and Regression Trees: Sınıflama ve Regresyon Ağaçları) yöntemi 1963 senesinde Sonquist ve Morgan, CART yöntem bilimini oluşturan ilk adımları atmışlardır (Da Rosa, Veiga, & Medeiros, 2008). Bu algoritma; 1984 yılında Breiman, Friedman, Olshen ve Stone tarafından geliştirilmiştir (Pehlivan G., 2006). Sınıflama ve Regresyon Ağaçları Tekniği (CART) sürekli veya kategorik bağımlı değişkenlerden oluşabilen bir tekniktir ve parametrik olmayan istatistiksel bir metodtur. Kategorik bağımlı değişkenlerde sınıflama ağacı, sürekli bağımlı değişkenlerde ise regresyon ağacı şeklinde tanımlanır.

CART yöntemi, veri setinin çok karmaşık olduğu durumlarda dahi bağımlı değişkeni etkileyen

değişkenleri ve bu değişkenlerin modeldeki önemini basit bir ağaç yapısı ile görsel olarak sunabilmektedir (Temel GO., 2005). 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 değerin hangi sınıfa düşeceğini belirlemektir. Bu amaçla, hangi değişkenlerin ya da değişkenler arası etkileşimin en iyi sonucu tahmin etmek için gerekli olduğu belirlenir (Yohannes et al., 1999).

Yöntemin en büyük avantajı parametrik olmayan bir yöntem olmasıdır. Bağımsız değişkenlerin sıralı, kategorik veya sürekli olması durumunda da rahatlıkla hesap yapılabilir. Analizi yapacak olan kişiye yöntem sıralaması üzerinde düzeltme yapma imkanı sağlar (Lewis RJ., 2000). Bağımlı ve bağımsız değişken fark etmeksizin eksik, kayıp ya da ekstrem değerlerin olumsuz etkilerinin gözlemlenmediği bir yöntem olduğundan kullanımı kolaydır.

CART, çoğunlukla tıp alanında teşhis ve öngörü amacıyla, karar teorisinde ve botanik alanında kullanılmaktadır.

Sonuç olarak CART yöntemi bütün başlangıç veri setini içinde barındıran kök düğümden başlayarak her düğümü iki küçük düğüme böler ve ikili ağaçlar oluşturur. CART algoritmasının oluşum mekanizması, düğümdeki homejenliği en üst düzeye çıkarabilmek için çalışır. Bir düğümün içinde homojen bir alt kümenin bulunması düğümün safsızlığının bir göstergesi olur. Yani bir uç düğüm her durumda bağımlı değişken için aynı değere sahipse bölünme yapmaz, çünkü artık saf bir düğümdür (Antipo et al., 2010).

CHAID ALGORİTMASI

Otomatik Ki-Kare Etkileşim Belirleme (CHAID) (Chi-Squared Automatic Interaction Detector) Analizi, kategorik bağımlı değişkenler için oluşturulmuş olup, AID analizinin bir uzantısı olarak kabul edilir. Otomatik Ki-Kare Etkileşim Belirleme Analizinin asıl amacı; veriyi daha homojen bir şekilde birden çok alt gruba bölmektir. Devasa bir veri kümesinin homojen bir alt gruba indirgenmesi; bağımlı değişkeni mümkün olduğunca tutumlu olarak açıklayan diğer değişkenleri ve bunlarla ilgili verileri meydana getirmek anlamına gelir. Otomatik Ki-Kare Etkileşim Belirleme Analizi; kategorik değişkenlerle ilgili veri kümesini, bağımlı değişkeni en iyi açıklayan türde detaylı ve homojen alt gruplara böler. Bu alt grupların en önemli özelliği, tahmin edici olmalarıdır. Seçilen tahmin edici gruplar; daha sonra yapılacak ileri analizlerde bağımlı değişkenin tahmininde kullanılır. Otomatik Ki-Kare Etkileşim Belirleme Analizi, regresyon analizlerinde kullanılabileceği gibi karar ağaçlarının oluşturulmasında da kullanılabilir. Değişkenler arasındaki ilişki lineer yapıdan daha karmaşık ise veride gizli olan bu ilişkiyi bulmak için verinin belli kısımlarını eleme yöntemi olan CHAID

kullanılır. "Ki-kare" ismini de almasının nedeni algoritmasında birçok çapraz tablonun kullanılması ve istatistiksel önem oranları ile çalışmasıdır (Hoare R., 2004).

CHAID analizinde bağımsız değişkenlerin herbiri için en iyi dallanma hesaplanır. Daha sonra bağımsız değişkenler en iyisi seçilene kadar karşılaştırılır. Seçilen en iyi bağımsız değişkene göre tekrar dallanma işlemi yapılır. Her bir bağımsız değişken kategorilerinin en anlamlı şekilde dallanma işlemi gerçekleştirildikten sonra bağımlı değişkene göre kontenjans tabloları oluşturulurak Bonferroni p değerleri ile χ^2 istatistikleri hesaplanır. Hesaplanan istatistikler doğrultusunda önem derecesine göre kontenjans tabloları şekillenmiş olur. Buradan da anlaşıldığı üzere CHAID analizi ki- kare istatistiklerini, Bonferronni yaklaşımını ve kategori birleştirme algoritmalarını kullanarak araştırmacının ağaç diyagramı ile en iyi açıklayıcı değişkenleri ve bağımlı değişken ile olan etkileşimleri elde etmesine olanak sağlar (Hoare R., 2004).

Bağımlı değişken kategori sayısı $d \geq 2$ olsun. Analiz edilecek olan belirli bir açıklayıcı değişken $c \geq 2$ sayıda kategoriye sahip olsun. Analizdeki amaç, c x d kontenjans tablosunu açıklayıcı değişkenindeki uygun kategorileri birleştirme yolu ile en anlamlı jxd tablosuna indirgemektir. İlk olarak Tj (i) istatistiği hesaplanır. Tj(i) j x d tablosunu oluşturmadaki i. metot için χ^2 istatistiğidir. (j: 2, 3, 4....., c; i'nin değişim aralığı açıklayıcı değişkenin tipine bağlıdır.) Tj*=maxi Tj(i) ise en iyi j x d tablo için, χ^2 istatistiği elde edilir (Kass GV., 1980).

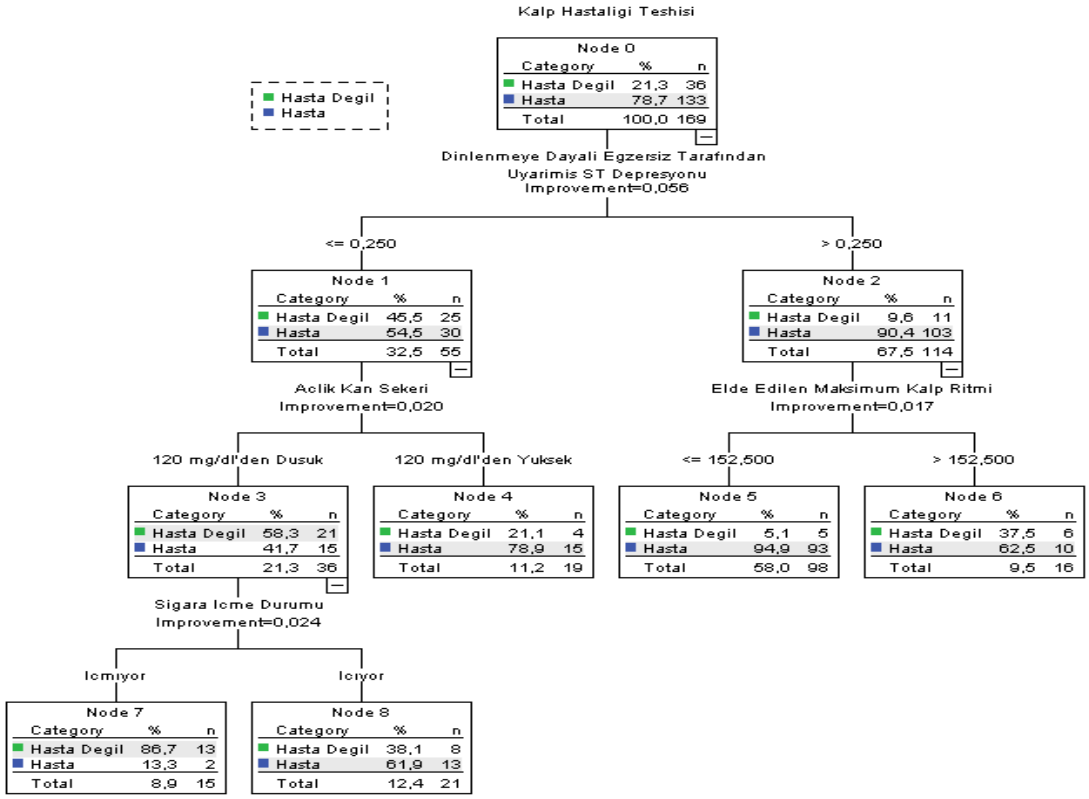
2. UYGULAMA

Bu çalışmada kalp hastalıklarına etki eden faktörlerin kendi aralarındaki ilişkileriyle, bağımlı değişken ve bağımsız değişkenleri arasındaki ilişki CART ve CHAID algoritmaları kullanılarak ayrı ayrı değerlendirilmiştir.

Çalışmada kalp rahatsızlığı durumu bağımlı değişken ve diğer 38 değişken ise bağımsız değişken olarak analize dahil edilmiştir. İlk önce CART algoritması ve ardından CHAID algoritması ile sınıflandırma ağaçları oluşturulmuş ve sonuçlar yorumlanmıştır.

Sınıflandırma ağacında öncelikle en uygun ağacın hangi ağaç olduğuna karar vermek gerekmektedir. Her iki algoritma için de uygun ağacı belirleyebilmek adına üç farklı deneme yapılmıştır. Bu çalışmada ilk olarak veri setinin % 70'i training sample ve %30'u da test sample olarak alınmıştır. İkinci olarak ise; veri setinin %50'si training sample %50'si de test sample olarak alınmıştır. Son olarak; veri setinin %30'u training sample ve %70'i de test sample olarak alınmıştır. 3 faklı oranla oluşturulan modellerde farklı düğüm sayıları ile uygun ağaç belirlenmeye çalışılmıştır. En yüksek doğru sınıflama oranına sahip olan ağacın en uygun ağaç olduğuna karar verilmiştir.

CHAID algoritması ile oluşturulan en uygun ağacın; veri setinin %70'i training sample %30'u test sample olarak alındığı ağaç olduğu görülmüştür.

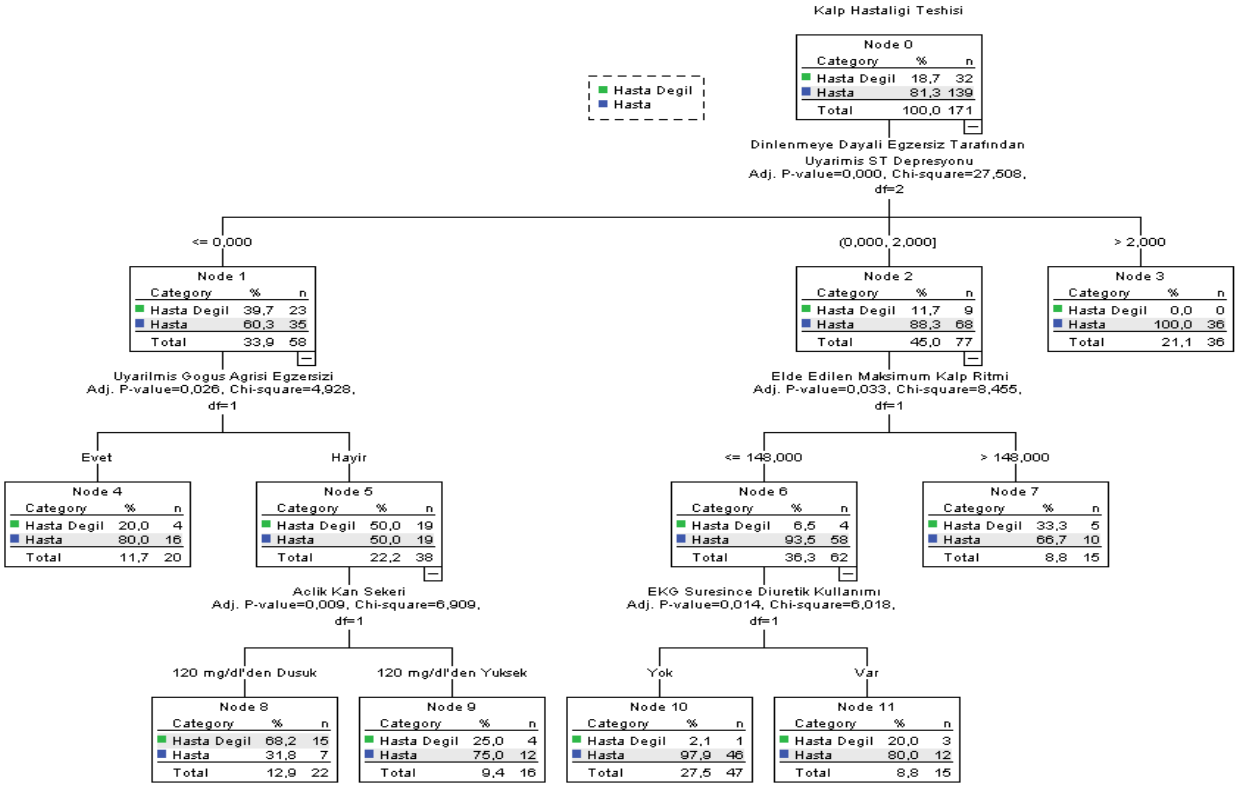


Şekil 1: CART Ağacı Diyagramı

Şekil 1 'de CART algoritması ile kalp rahatsızlığını etkileyen değişkenler incelenmiştir.

Şekil 1 incelendiğinde kalp hastalığını etkileyen değişkenler; dinlenmeye dayalı egzersiz tarafından uyarılmış ST depresyonu, açlık kanşekeri, kalp ritmi sayısı ve sigara içme durumudur. Çalışma sonucuna göre bu değişkenler arasında ilişki olduğu sptanmıştır. Kalp hastalıklarını etkileyen en önemli değişkenin ST depresyonu olduğu sonucuna varılmıştır. ST depresyon grafisi değeri 0.25 ten fazla olanların % 90.4 'ü hasta, 0,25 ten düşük olanların ise % 54.5 'inin hasta olduğu görülmüştür. ST depresyon değeri düştükçe, kalp rahatsızlığı olanların oranı da düşmektedir.

ST depresyon değişkenini açlık kan şekeri ve kalp ritmi değişkenin etkilediği görülmektedir. Kan şekeri oranı 120 mg/dl 'den düşük olanların %41,7 'sinin kalp rahatsızlığı da bulunmaktadır. Kan şekeri oranının 120mg/dl 'den yüksek olan kişilerin %78,9 'u kalp rahatsızlığı teşhisi konulmuşken, maksimum kalp ritmi, 152 'den düşük olanların % 94,9 'u kalp rahatsızlığı yaşarken, 152 'den yüksek olanlarda ise oranın % 62.5 'e düştüğü tespit edilmiştir. Sigara kullananlarda ise oran %61,9 'dur.



Şekil 2: CHAID Ağacı Diyagramı

Şekil 2 incelendiğinde; kalp hastalığını etkileyen değişkenler; Dinlenmeye dayalı egzersiz tarafından uyarılmış ST depresyonu, elde edilen maksimum kalp ritmi, uyarılmış gögüs ağrısı egzersizi, açlık kan şekeri ve EKG süresince diüretik (idrar söktürücü) kullanımı ve bu değişkenler arasında ilişki olduğu sonucuna varılmıştır.

Kalp hastalığını etkileyen en önemli değişkenin ST depresyonu (p=0.00, $\chi 2$=27.508, sd=2) olduğu sonucuna varılmıştır. Bu değişken üç düğüm olarak ayrılmıştır. Bu düğümlerden en önemlisi ST depresyonu değeri 2 'den yüksek olan kişilerin % 100 'ünün hasta olduğu görülmüştür. Bu sayı azaldıkça hasta sayısının da azaldığı tespit edilmiştir.

ST depresyonu değeri (0-2 arasında olan kişi), değişkeni elde edilen maksimum kalp ritmi değişkenini (p=0,033 $\chi 2$=8.455, sd=1) etkilemektedir. Maksimum kalp ritmi 148/dk 'dan az olan kişilerde kalp rahatsızlığı olan hasta oranı %93,5 olarak göstrilmiştir. 148 'den fazla olanlarda ise hasta oranı % 66,7'dir. Ayrıca maksimum kalp ritmi değişkeniyle EKG süresince diüretik kullanımı (p=0,014 $\chi 2$=6.018, sd=1) değişkeniyle ilişkili olduğu saptanmıştır. EKG süresince diüretik kullanmayanların %97,9 'nun hasta olduğu görülmüştür.

Sonuç

Bu çalışmada; University of California bünyesinde veri setlerini barındıran bir platformdan alınan ve kalp hastalığına etki eden 38 adet faktör kullanılmıştır. Bu faktörlerin değerlendirilmesi; Classification and Regression Tree (CART) ve Chi-Squared Automatic Interaction Detector (CHAID) algoritmaları kullanılarak oluşturulmuştur ve çıkan sonuçlar birbirleriyle karşılaşırılarak yorumlanmıştr.

Çalışmada kalp hastalığı bağımlı değişken olarak alınmıştır. Her iki algoritma ile oluşturulan ağaçlarda; kalp hastalığına etki eden bağımsız değişkenler gösterilmektedir. CART algoritması ile oluşturulan sınıflandırma ağacına göre ise kalp hastalığını etkileyen değişkenler; dinlenmeye dayalı egzersiz tarafından uyarılmış ST depresyonu, açlık kan şekeri, kalp ritmi ve sigara içme durumudur. CHAID algoritması ile oluşturulan ağaca göre kalp hastalığını etkileyen değişkenler; dinlenmeye dayalı egzersiz tarafından uyarılmış ST depresyonu, elde edilen maksimum kalp ritmi sayısı, uyarılmış gögüs ağrısı egzersizi, açlık kan şekeri ve EKG süresince diüretik kullanımıdır. Bu değişkenler arasında ilişki olduğu sonucuna varılmıştır.

Araştırmaya dahil edilen kişilerin %80 'inde kalp hastalığı olduğu görülmüştür. Her iki ağaçta da kalp hastalığını etkileyen en önemli değişkenin dinlenmeye dayalı egzersiz tarafından uyarılmış ST depresyonu olduğu saptanmıştır. ST depresyonu değeri arttıkça kalp rahatsızlığının da arttığı görülmüştür. Dinlenmeye dayalı egzersiz tarafından uyarılmış ST depresyonu değişkeni, CART algoritmasında elde edilen maksimum kalp ritmi değişkeninden etkilenirken, CHAID algoritmasında da aynı sonuca ulaşılmıştır. CART ve CHAID algoritmasında maksimum kalp ritmi düzeyleri düştükçe kalp hastalığı olanların oranıda yükselmiştir. CART algoritmasında bireylerin sigara kullanımına bakıldığında; sigara kullananların kalp hastalığına yakalanma olasılığının arttığı görülmüştür. CHAID algoritmasında ise EKG süresince diüretik kullanmayanlarda %97,9 oranında kalp rahatsızlığı görülmektedir.

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

Kaynaklar

Atılgan, E. (2011). Karayollarında Meydana Gelen Trafik Kazalarının Karar Ağaçları ve Birliktelik Analizi İle İncelenmesi. Yüksek Lisans Tezi. Hacettepe Üniversitesi İstatistik Anabilim Dalı.

F, H., & McCulloch, R. E. (2000). Hierarchical Priors for Bayesian CART Shrinkage. Statistics and Computing, 10(1), 17-24.

Cumhur M.(2014). Temel Anatomi. 4.Baskı.Ankara: ODTÜ Yayıncılık.

Da Rosa, J. C., Veiga, A., & Medeiros, M. C. (2008). Tree-Structured Smooth Transition Regression Models. Computational Statistics & Data Analysis, 52(5), 2469-2488.

Evgeny Antipov ve Elena Pokryshevskaya, "Applying CHAID for logistic regression diagnostics and classification accuracy improvement", Journal of Targeting, Measurement and Analysis for Marketing, 2010, Vol.18).

Fu, C. Y. (2004). Combining Loglinear Model with Classification and Regression Tree (Cart): An Application to Birth Data. Computational Statistics & Data Analysis, 45(4), 865-874.

Gülhan Orekici Temel, Handan Çamdeviren, Zeki Akkuş, "Sınıflama Ağaçları Yardımıyla Legs Syndrome (RLS) Hastalarına Tanı Koyma", İnönü Üni. Tıp Fak. Dergisi, Cilt.12, Sayı.2 (2005), s.111.

Haldun AKPINAR. (2017). DATA Veri madenciliği, Veri Analizi. 2.Baskı. İstanbul: Papatya Yayıncılık Eğitim.

Hoare, R., (2004), "Using CHAID for Classification Problems", New Zealand Statistical Association, New Zealand.

Kass, G.V.,(1980), "An Exploratory Technique for Investigating Large Quantities of Categorical Data", Applied Statistic, 1980,29(2):119-127.

Lewis, R. J. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. Annual Meeting of the Society for Academic Emergency Medicine, (s. 1-14). Kaliforniya

Özekes S. (2002). Veri Madenciliği Uygulaması. Yüksek Lisans Tezi, Marmara Üni. Fen Bilimleri Enstitüsü.

Pehlivan Gamze. "Chaid Analizi ve Bir Uygulama", Yayınlanmamış Yüksek lisans Tezi. Yıldız Teknik Üniversitesi, FBE, 2006.

Quinlan, J.R., 1986. Induction of Decision Trees. Journal of Machine Learning, Cilt 1, s 81-106.

Silahtaroğlu G. Veri Madenciliği (Kavram ve Algoritmaları). 2. Basım,

İstanbul: Papatya Yayıncılık Eğitim, 2013.

Ulusoy, G., 2013. Karar Ağacı Analizi ile AB GeniĢleme Kriterlerinin Değerlendirilmesi. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Ġstatistik Bilim Dalı. Yüksek Lisans Tezi.

Yohannes, Y., & Hoddinott, J. (1999). Classification and Regression Trees: an Introduction. International Food Policy Research Institute.

Yusuf, S.;Reddy, S.; Ounpuu, S.; Anand, S. (2001).Global burden of cardiovascular diseases Part 1:General considerations, the epidemiologic transition, risk factors and impact of urbanization. Clinical Cardiology: New Frontiers, 104:2746-2753).