



**KEÇİLERE AİT KİMİ ÖZELLİKLERİN CANLI
AĞIRLIK ÜZERİNDEKİ ETKİLERİNİ BELİRLEMEK
AMACIYLA KULLANILAN VERİ MADENCİLİĞİ
ALGORİTMALARININ (CART, CHAID VE MARS)
KARŞILAŞTIRILMASI**

Harun SEVGENLER
Yüksek Lisans Tezi

ZOOTEKNİ ANABİLİM DALI
Danışman: Prof. Dr. Ecevit EYDURAN
2019

**T.C.
IĞDIR ÜNİVERSİTESİ**

FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

**KEÇİLERE AİT KİMİ ÖZELLİKLERİN CANLI AĞIRLIK ÜZERİNDEKİ
ETKİLERİNİ BELİRLEMEK AMACIYLA KULLANILAN VERİ MADENCİLİĞİ
ALGORİTMALARININ (CART, CHAID VE MARS) KARŞILAŞTIRILMASI**

Harun SEVGENLER

ZOOTEKNİ ANABİLİM DALI

IĞDIR

2019

Her hakkı saklıdır

Prof. Dr. Ecevit EYDURAN danışmanlığında Harun SEVGENLER tarafından hazırlanan bu çalışma 28/01/2019 tarihinde aşağıdaki jüri üyeleri tarafından Zootekni Anabilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. Ecevit EYDURANİmza:

Üye:İmza:

Üye:İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulunun / /2019 tarih ve 2019/ sayılı kararı ile onaylanmıştır.

(imza)

.....

Doç. Dr. Süleyman TEMEL

Enstitü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.



Harun SEVGENLER

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZET

KEÇİLERE AİT KİMİ ÖZELLİKLERİN CANLI AĞIRLIK ÜZERİNDEKİ ETKİLERİNİ BELİRLEMEK AMACIYLA KULLANILAN VERİ MADENCİLİĞİ ALGORİTMALARININ (CART, CHAID VE MARS) KARŞILAŞTIRILMASI

SEVGENLER, Harun

Yüksek Lisans Tezi, Zootekni Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ecevit EYDURAN

Ocak 2019, 57 sayfa

Bu çalışmanın amaçları, hayvancılıkla ilgili bir veri setinden yola çıkarak CART, CHAID, Exhaustive CHAID ve MARS veri madenciliği algoritmalarının tahmin performanslarını karşılaştırmalı olarak incelemek ve bu algoritmalara ilişkin uyum iyiliği ölçütlerinin hesaplandığı R script dosyalarını oluşturmaktır. Bu amaca ulaşmak için sürekli bir değişkeni (canlı ağırlık) tahmin etmek amacıyla Teddy ırkı 164 baş keçiye ait bazı morfolojik özellikler bağımsız değişken olarak kullanılmıştır. İncelenen algoritmaların tahmin performanslarını ölçmek için, bağımlı değişken bakımından gerçek ve tahmin edilen değerler arasındaki Pearson korelasyon katsayısı (r), belirleme katsayısı, (R^2), düzeltilmiş determinasyon katsayısı, (Düz. R^2) Akaike bilgi ölçütü, (AIC), düzeltilmiş Akaike bilgi ölçütü, (Düz.AIC), hata kareler ortalamasının karekökü, (HKOK), Ortalama hata, (OH), Ortalama mutlak sapma, (OMS), standart sapma oranı, (SS Oran), Global nispi yaklaşık hata, (NYH), varyasyon katsayısı (VK %) ve Ortalama mutlak yüzde hata, (OMYH) uyum iyiliği ölçütleri kullanılmıştır. Ağaç yapısına dayalı CART, CHAID ve Exhaustive CHAID algoritmalarında en yüksek tahmin doğruluğuna ulaşmak için ebeveyn ve çocuk düğümlerindeki minimum hayvan sayısı 4:2 olarak ayarlanmıştır. MARS algoritmasında en iyi tahmin doğruluğu, 4. derece interaksiyon ve 45 terim tanımlamaları ile elde edilmiştir. Kantitatif bir özelliğin (canlı ağırlık) tahmin edilmesi kapsamında en iyi tahmin performansı MARS algoritması ile elde edilmiştir. Tahmin performansı bakımından MARS algoritmasını, Exhaustive CHAID > CHAID > CART algoritmaları izlemiştir.

Sonuç olarak elde edilen sonuçlar ışığında gerek çalışılan ırk standartlarının daha iyi tanımlanması gerekse önemli bağımsız değişkenlerin tespit edilmesi bakımından MARS algoritmasının kullanılmasının daha iyi bir seçenek olabileceği sonucuna varılmıştır.

Anahtar kelimeler: Veri madenciliği, CART, CHAID, Exhaustive CHAID, MARS, Regresyon ağaçları.

ABSTRACT

COMPARISON OF DATA MINING ALGORITHMS (CART, CHAID AND MARS) USED TO DETERMINE THE EFFECTS OF SOME CHARACTERISTICS ON BODY WEIGHT IN GOATS

Harun SEVGENLER

Master Thesis, Animal Science Main Discipline

Thesis Adviser: Prof. Dr. Ecevit EYDURAN

January 2019, 57 pages

The main aims of this study were to comparatively examine predictive performances of the CART, CHAID, Exhaustive CHAID and MARS data mining algorithms on the basis of a data set relevant to animal husbandry, and create R script files for estimating their goodness of fit criteria. To achieve this purpose, several morphological traits measured from 164 Teddy goats with the aim of predicting a continuous response variable (live body weight) were used as explanatory variables. To measure predictive performances of the examined algorithms, goodness of fit criteria i.e. Pearson correlation coefficient in the response variable between actual and predicted values (r), coefficient of determination (R^2), adjusted coefficient of determination ($Adj.R^2$), Akaike information criterion (AIC), the corrected Akaike information criterion (AICc), root-mean-square error (RMSE), mean error (ME), mean absolute deviation (MAD), standard deviation ratio (SDratio), global relative approximation error (RAE), coefficient of variation (CV%) and mean absolute percentage error, MAPE) were employed. To ensure the highest predictive accuracy of tree-based CART, CHAID and Exhaustive CHAID data mining algorithms, minimum animal numbers in parent and child nodes were set at 4:2. The highest predictive accuracy for the MARS algorithm was obtained with describing 4th interaction order and 45 terms. With the scope of predicting a quantitative trait (live body weight), the best performance was obtained with MARS algorithm, followed by the Exhaustive CHAID, CHAID and CART data mining algorithms.

It was concluded in the light of the obtained results that use of MARS algorithm may be a better choice in both better describing the studied breed standards and determining significant predictors.

Keywords: Data mining, CART, CHAID, Exhaustive CHAID, MARS, Regression trees.

ÖNSÖZ ve TEŞEKKÜR

Regresyon analizi kapsamında bazı temel varsayımların sağlanmaması birtakım problemlerin oluşmasına sebebiyet vermektedir. Bu problemlerin giderilmesi bakımından daha etkin ve yüksek tahmin doğruluğu üreten veri madenciliği algoritmaları kullanılabilir. İsbetli kararlar verilmesini sağlayan bu algoritmalar, sınıflayıcı (nominal), sıralı (ordinal) ve sürekli (continuous) değişkenler için rahatlıkla uygulanabilirler. CART, CHAID, Exhaustive CHAID, QUEST, C5.0 ve MARS bu algoritmalara örnek verilebilir. Bahsedilen bu algoritmalarından ilk üçü, homojen alt gruplar üretmek amacıyla hem regresyon ağacı oluşturmada hemde sınıflandırma ağacı oluşturmada yaygın olarak kullanılmaktadır. Ancak, CART ve CHAID algoritmalarının aksine tahmin denklemleri üreten MARS algoritması hem bir ya da birden fazla bağımlı değişkenin eş zamanlı olarak etkin bir şekilde analiz edilmesini hem de bağımlı değişkenleri etkileyen değişkenlerden elde edilen yüksek dereceli interaksiyon etkilerinin tespit edilmesini sağlayabilmektedir. Bağımsız değişkenlerin dağılımına ilişkin herhangi bir varsayıma gerek duymayan MARS algoritması gerek yapay sinir ağları algoritmalarına gerekse yanıt yüzeyi yöntemine alternatif olabilmektedir. Veri madenciliği algoritmaları kapsamında, SPSS ve STATISTICA gibi yüksek maliyetli yazılımlar uyum iyiliği ölçütlerine ilişkin kapsamlı bir çıktı üretememektedir. Bu nedenle bu tez çalışması ile ele alınan non-parametrik CART, CHAID, Exhaustive CHAID ve MARS algoritmalarına ilişkin uyum iyiliği ölçütlerini hesaplanmasını sağlayan R script dosyaları literatüre kazandırılmıştır.

Tez çalışmamın her aşamasında yakın ilgi ve desteğini gördüğüm ayrıca bilimsel kimliğinin yanında kişiliğini örnek aldığım danışman hocam Sayın Prof. Dr. Ecevit EYDURAN'a en derin sevgi, saygı ve teşekkürlerimi sunarım.

Bu tez çalışmamda verilerinin kullanılmasına izin veren Dr. Abdul WAHEED'e teşekkürü bir borç bilirim.

Harun SEVGENLER

Şubat, 2019

İÇİNDEKİLER

	Sayfa No
ÖZET.....	i
ABSTRACT.....	ii
ÖNSÖZ ve TEŞEKKÜR.....	iii
İÇİNDEKİLER	iv
SİMGELER ve KISALTMALAR DİZİNİ.....	v
ŞEKİLLER DİZİNİ.....	vii
ÇİZELGELER DİZİNİ	viii
1. GİRİŞ	1
2. KAYNAK ÖZETLERİ.....	4
3. MATERYAL ve METOT.....	15
3.1. Materyal	15
3.2. Metot	16
4. BULGULAR ve TARTIŞMA.....	22
4.1. Algoritmaların Karşılaştırılması	22
4.2. Exhaustive CHAID algoritmasına ilişkin sonuçlar	25
4.3. CHAID algoritmasına ilişkin sonuçlar.....	28
4.4. CART algoritmasına ilişkin sonuçlar.....	31
4.5. MARS algoritmasına ilişkin sonuçları.....	33
5. SONUÇ ve ÖNERİLER.....	43
KAYNAKLAR	46
EKLER	
EK 1	53
EK 2	56
ÖZGEÇMİŞ	58

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler

%	Yüzde
*	Çarpım
$M(\lambda)$	λ terimlerini içeren modelin karmaşıklığına ilişkin ceza fonksiyonu
y_{ip}	Bağımlı değişkenin tahmin değeri
\hat{y}	Bağımlı değişkenin tahmin değeri
y_i	Bağımlı değişkene ait gözlem değeri
cm	Santimetre
Düz. R^2	Düzeltilmiş belirleme katsayısı
Düz.P	Düzeltilmiş önem düzeyi
K	Model parametrelerinin sayısı
kg	Kilogram
n	Veri setindeki havyan sayısı
P	Önem düzeyi (olasılık)
r	Pearson korelasyon katsayısı
R^2	Belirleme katsayısı
s_d	Bağımlı değişkenin standart sapması
s_m	Model hata terimlerinin standart sapması

Kısaltmalar

AIC	Akaike's information criterion (Akaike bilgi ölçütü)
AIC_c	Corrected Akaike's information criterion (Düzeltilmiş Akaike bilgi ölçütü)
CART	Classification and Regression Tree
CHAID	Chi-Square Automatic Interaction Detector
EKK	En küçük kareler
GCV	Global relative approximation error (Global nispi yaklaşık hata)
MAD	Mean absolute deviation (Ortalama mutlak sapma)
MAPE	Mean absolute percentage error (Ortalama mutlak yüzde hata)
MARS	Multivariate Adaptive Regression Splines
ME	Mean error (Ortalama hata)
MLP	Multilayer Perceptron
OLS	Ordinary Least Squares
PCA	Principal component analysis
RAE	Generalized cross-validation (Genelleştirilmiş çapraz geçerlilik)
RBF	Radial Basis Function
RMSE	Root-mean-square error (Hata kareler ortalamasının karekökü)
SD_{ratio}	Standard deviation ratio (Standart sapma oranı)
SPSS	Statistical Package For Social Sciences

EKİLLER DİZİNİ

	Sayfa No
Şekil 4.1. MARS algoritması için uyum grafiği	23
Şekil 4.2. Exhaustive CHAID algoritması için uyum grafiği	23
Şekil 4.3. CHAID için uyum grafiği	24
Şekil 4.4. CART algoritması için uyum grafiği	24
Şekil 4.5. Exhaustive CHAID algoritması ile oluşturulan ağaç diyagramı	27
Şekil 4.6. CHAID algoritması ile oluşturulan ağaç diyagramı	30
Şekil 4.7. CART algoritması ile oluşturulan ağaç diyagramı	32

ÇİZELGELER DİZİNİ

Sayfa No

Çizelge 3.1. Sürekli bağımsız değişkenlerin tanıtıcı istatistikleri.....	15
Çizelge 4.1. Algoritmalara ilişkin uyum iyiliği ölçütleri.....	22
Çizelge 4.2. MARS tahmin denklemine ait terimler ve katsayıları	34
Çizelge 4.3. MARS maskelenmiş tahmin denklemi II'e ait terimler ve katsayıları	35
Çizelge 4.4. MARS maskelenmiş tahmin denklemi III'e ait terimler ve katsayıları	36
Çizelge 4.5. MARS maskelenmiş tahmin denklemi IV'e ait terimler ve katsayıları.....	36
Çizelge 4.6. Canlı ağırlığı tahmin etmek amacıyla eski yıllarda yapılan bazı çalışmalar ...	40

1. GİRİŞ

Hayvancılık alanında verim özelliğini etkileyen faktörlerin ve bu faktörlere ilişkin interaksiyon etkilerinin daha doğru ve güvenilir bir şekilde ortaya konulması yeni ıslah stratejilerinin geliştirilmesi ve çalışılan hayvan ırklarının ırk standartlarının belirlenmesi açısından faydalı olacaktır. Üzerinde durulan verim özelliği bakımından değişkenler arasındaki ilişkilerin etkin bir biçimde tanımlanması, güçlü algoritmaların kullanılması ile birlikte bağımsız değişkenlerin doğru seçilmesine bağlıdır. Bu bakımdan, sınıflayıcı (nominal), sıralı (ordinal) ve sürekli (continuous) değişkenlerin analizinde veri madenciliği yani yapay zekâ algoritmalarının kullanımı ön plana çıkmaktadır. Bu algoritmaların kullanımı, bağımlı değişken bakımından en iyi sonucu veren bağımsız değişken kombinasyonlarının belirlenmesinde avantaj sağlamaktadır.

Sürekli bağımlı değişkeni (continuous response variable) etkileyen bağımsız değişken kombinasyonlarının tespit edilmesi regresyon problemleri kapsamında değerlendirilir. Bu bağlamda, en küçük kareler (EKK) metodunun güvenilirliği bazı temel regresyon varsayımların sağlanıp sağlanamamasına ve bağımsız değişkenler arasındaki güçlü korelasyonlardan kaynaklanan çoklu bağlantılılık (multicollinearity) probleminin oluşup oluşmamasına bağlıdır (Khan *et al.*, 2014). Çoklu bağlantılılık durumunda regresyon katsayılarına ait parametre tahminleri yanlıdır. Çoklu bağlantılılığı gidermek için kullanılan en popüler yaklaşım Ridge regresyon analiz (Ridge regression analysis) metodudur (Eyduran *et al.*, 2013). Çoklu bağlantılılığı gidermek amacıyla, çoklu regresyon analizi (multiple linear regression analysis) için açıklayıcı faktör analizi (explanatory factor analysis) ve temel bileşenler analizi (principal component analysis, PCA) gibi çok değişkenli analiz (multivariate analysis) metotları ile birbirinden bağımsız latent değişkenler (uncorrelated latent variables) elde edilebilir (Eyduran *et al.*, 2009; Khan *et al.*, 2014). Ancak, söz konusu bu iki çok değişkenli analiz tekniğinin etkinliği örnek genişliğinin büyüklüğüne ve korelasyon matrisinin birim matris olmamasına bağlıdır.

Son yıllarda bilgisayar teknolojisindeki gelişmelere paralel olarak çok değişkenli karmaşık veri setlerinin istatistiksel analizlerinin yapılması için oldukça etkin

algoritmaların kullanımı yaygınlaşmıştır (Aytekin *et al.*, 2018). Hayvan ıslahı açısından et, süt ve yumurta gibi ekonomik öneme sahip kantitatif özellikleri etkileyen faktörlerin doğru tespit edilebilmesi ve seleksiyon stratejilerinin geliştirilebilmesi için etkin istatistiksel metotların kullanılması önem arz etmektedir (Koc *et al.*, 2017). Bu kapsamda, temel istatistiksel varsayımların yerine gelmediği durumlarda oldukça iyi performans gösteren CART (Classification and Regression Tree), CHAID (Chi-Square Automatic Interaction Detector), Exhaustive CHAID, MLP (Multilayer Perceptron) ve RBF (Radial Basis Function) gibi veri madenciliği (yapay zeka) algoritmalarının kullanılması doğru bir yaklaşım olacaktır (Celik *et al.*, 2017; Eydurun *et al.*, 2017). Bu algoritmalar arasında ağaç yapısına dayalı CART, CHAID ve Exhaustive CHAID algoritmaları heterojen yapılı veri setinden mümkün olduğunca homojen alt gruplar oluşturana kadar yinelemeli olarak bölünmeye devam eden bir ağaç yapısı oluşturur. Bu kapsamda, sürekli bağımlı değişken için oluşturulan ağaca regresyon ağacı adı verilmektedir. CART algoritmasının bir modifikasyonu olan MARS veri madenciliği algoritması, değişkenlerin dağılımına ilişkin herhangi bir varsayım gerektirmemesinin yanında bağımlı ve bağımsız değişkenlerin arasında fonksiyonel bir varsayım gerektirmemesi bakımından öne çıkmaktadır. Genelde kullanılan doğrusal ve doğrusal olmayan regresyon modellerinin aksine, bir non-parametrik regresyon metodu olarak da bilinen MARS algoritması bağımsız değişkenin farklı aralık değerleri için farklı katsayılar türeterek ve interaksiyon terimlerini de hesaba katarak yüksek tahmin performanslı modeller geliştirilmesine olanak sağlar. MARS algoritması, tahmin performansını geliştirmek açısından yinelemeli ayırma metodunun ve adimsal doğrusal regresyonun genelleştirilmiş bir hali olarak düşünülebilir. MARS, bağımlı ve bağımsız değişkenler arasındaki doğrusal olmayan ilişkileri doğrusal yapıya dönüştürme amacıyla uygun dönüştürme teknikleri kullanır (Temel *ve ark.*, 2010). MARS, bağımsız değişkenlerin farklı değer aralıklarına karşılık gelen temel fonksiyonları kullanarak esnek bir regresyon modeli oluşturur. Ayrıca, yanıt yüzeyi yöntemine (response surface method) alternatif olan MARS algoritması ile birden fazla kantitatif özelliğin eş zamanlı analizinde rahatlıkla uygulanabilir (Eydurun, 2016).

Hayvancılık alanında ağaç yapısına dayalı CART (Yakubu *et al.*, 2012; Ali *et al.*, 2015; Celik and Yilmaz, 2017; Celik *et al.*, 2017; Eyduran *et al.*, 2017a; Karabacak *et al.*, 2017; Koc *et al.*, 2017; Celik and Yilmaz, 2018; Gevrekci and Takma, 2018; Olfaz *et al.*, 2019), CHAID (Doğan, 2003; Eyduran *et al.*, 2008; Bakır *et al.*, 2009; Mendeş *et al.*, 2009; Bakır *et al.*, 2010; Oruçoğlu, 2011; Mohammad *et al.*, 2012; Akşahan, 2015; Ali *et al.*, 2015; Eyduran, 2016; Orhan *et al.*, 2016; Celik and Yilmaz, 2017; Celik *et al.*, 2017; Duru *et al.*, 2017; Eyduran *et al.*, 2017a; Karabacak *et al.*, 2017; Koc *et al.*, 2017; Tahtalı ve Berberoğlu, 2017; Aksoy *et al.*, 2018a; Gevrekci and Takma, 2018; Olfaz *et al.*, 2019) ve Exhaustive CHAID (Eyduran *et al.*, 2013b; Khan *et al.*, 2014; Celik and Yilmaz, 2017; Celik *et al.*, 2017; Karabacak *et al.*, 2017; Koc *et al.*, 2017) algoritmalarının aksine, sürekli değişkenin tahmin edilmesi kapsamında MARS (Multivariate Adaptive Regression Splines) algoritmasının az kullanıldığı bildirilmiştir (Celik *et al.*, 2017; Eyduran *et al.*, 2017b; Celik and Yilmaz, 2018; Ertürk, 2018; Ertürk *et al.*, 2018; Zaborski *et al.*, 2018). Bu nedenle, bu tez çalışması ile CART, CHAID ve MARS veri madenciliği algoritmalarının tahmin performanslarının karşılaştırılması ve incelenen hayvan ırkına ilişkin ırk standartlarının tespit edilmesi literatürde önemli bir boşluk doldurulacaktır. Ayrıca, literatürdeki çalışmalar incelendiğinde, SPSS ve STATİSTİCA gibi istatistik programlarında uyum iyiliği ölçütlerinin hesaplanması ile ilgili detaylar yetersizdir. Bu eksikliği gidermek amacıyla yapılan bu çalışmada, söz konusu algoritmaların tahmin performanslarının karşılaştırılması açısından yazılacak R script dosyaları literatüre kazandırılmış olacaktır.

2. KAYNAK ÖZETLERİ

Doğan (2003), çalışılan verim özelliğini etkileyen faktörlerin hangi seviyesinin etkili olduğunu bulmak için kullanılan CHAID analiz yönteminin, bir popülasyonu, bağımlı değişkendeki varyasyonu gruplar içi minimum ve gruplar arası maksimum olacak şekilde farklı alt gruplara tekrarlı olarak ayıran bir teknik olduğundan söz etmiştir.

Yakubu (2012) yaptığı çalışmada, Uda koyunlarında farklı vücut ölçüleri kullanarak canlı ağırlığı tahmin etmek amacıyla CART regresyon ağacı algoritmasını uygulamışlardır. Canlı ağırlığın tahmininde en önemli bağımsız değişkenlerin; göğüs çevresi, omuz genişliği, sağrı genişliği, vücut ve yüz uzunluğu olduğu belirlenmiştir. Bu algoritma ile elde edilen regresyon ağacı incelendiğinde, göğüs çevresi 94,05 cm'den geniş ve yüz uzunluğu 20,85 cm'den daha uzun olan hayvanların daha ağır olacağı belirlenmiştir. Sonuç olarak, araştırmacılar elde edilen bu sonucun, Uda koyunlarının bakım-idare, seleksiyon ve ıslah çalışmaları açısından önemli olabileceğini belirtmişlerdir.

Grzesiak *et al.* (2003), kısmi laktasyon kayıtlarını kullanarak 305 günlük süt verimini tahmin etmişlerdir. Araştırmacılar, süt verimi tahmininde, yapay sinir ağları (Artificial Neural Networks) ve çoklu regresyon (Multiple Linear Regression) algoritmalarını kullanmışlardır. Algoritmaların karşılaştırılmasında, hata kareler ortalamasının karekökü, standart sapma oranı, nispi ortalama hata ve düzeltilmiş belirleme katsayısı gibi uyum iyiliği ölçütleri kullanılmıştır. Araştırma sonucunda, 305 günlük süt verimi tahmin edilmesinde yapay sinir ağları algoritması kullanımının daha iyi olabileceği saptanmıştır.

Grzesiak *et al.* (2006), 305 günlük süt verimini tahmin etmek amacıyla yapay sinir ağları ve Wood metotlarını kullanmışlardır. Bu çalışmada en iyi modelin belirlenmesinde, nispi yaklaşık hatalar (NYH) ve hata kareler ortalamasının karekökü (HKOK) ve belirleme katsayısı (R^2) uyum iyiliği ölçütleri esas alınmıştır. Çalışmanın sonucunda, Wood metoduna nazaran yapay sinir ağları kullanımı önerilmiştir.

Eyduran *et al.* (2008), Norduz ve Karakaş kuzularının doğum ağırlığı üzerinde etkili olan faktörleri belirlemek amacıyla CHAID regresyon ağacı algoritmasını

kullanmışlardır. Regresyon ağacı incelendiğinde, doğum ağırlığını birinci dereceden doğum tipinin etkilediği görülmektedir (Düz. $P<0.01$). Ayrıca, doğum ağırlığını ikinci dereceden ırk ve cinsiyet etkilerken, üçüncü dereceden ise ana yaşının etkilediği tespit edilmiştir (Düz. $P<0,01$).

Bakır *et al.* (2009), CHAID algoritması ile Brown Swiss ırkı sığırlarda ergin çağ süt verimini tahmin etmek amacıyla bağımsız değişken (ineğin yaşı, ilk tohumlama yaşı, laktasyon sırası, laktasyon süresi, kuruda kalma süresi, ilk buzağılama yaşı, buzağılama mevsimi, doğum tipi ve cinsiyet) olarak kullanmışlardır. CHAID regresyon ağacı incelendiğinde, laktasyon süresinin ergin çağ süt verimi üzerinde birinci derecede etkili olduğu, cinsiyet ve ineğin yaşının ikinci derecede etkili olduğu ve ilk tohumlama yaşı ve kuruda kalma süresinin ise ergin süt verimini etkileyen önemli değişkenlerden olduğu belirlenmiştir (Düz. $P<0,001$).

Mendeş and Akkartal (2009), 224 adet Ross-308 etçi piliçlerin kesim ağırlığı üzerinde etkili olan değişkenleri belirlemek için CART regresyon ağacı algoritmasını kullanmışlardır. Araştırmacılar, incelenen bağımsız değişkenler arasında sadece canlı ağırlık, göğüs kemiği uzunluğu, bacak genişliği ve göğüs çevresinin kesim ağırlığı üzerinde etkili olduğunu belirtmişlerdir.

Bakır *et al.* (2010) değişik çiftliklerde yetiştirilen Holştayn ırkı sığırlar üzerinde yaptıkları çalışmada kuruda kalma süresi, laktasyon sırası, işletme, buzağılama mevsimi ve yaşı gibi çevre faktörlerinin 305 günlük süt verimi üzerindeki etkilerini tespit etmek amacıyla CHAID algoritmasını kullanmışlardır. Araştırmada, kuruda kalma süresi (Düz. $P<0,01$) ve laktasyon sırasının (Düz. $P<0,01$), 305 günlük süt verimi üzerinde birinci derecede etkili faktörler olduğu; işletme (Düz. $P<0,01$) ve buzağılama mevsimi (Düz. $P<0,01$) faktörlerinin ikinci derecede etkili faktörler olduğu ve buzağılama yaşı faktörünün ($P<0,01$) ise üçüncü derecede etkili olduğu saptanmıştır.

Oruçoğlu (2011) Holştayn ırkı ineklerin 305 günlük süt verimini etkileyen faktörleri belirlemek amacıyla CHAID veri madenciliği algoritmasını kullanmıştır. Araştırmacı, farklı ebeveyn ve çocuk düğüm oranları için CHAID algoritması ile oluşturulan regresyon ağacı analizinin tahmin performansının düşük olduğunu tespit etmiştir.

Mohammad *et al.* (2012), Pakistan’da yetiştirilen Mengali, Balochi, Harnai, Beverigh ve Rakshani ırkları üzerinde yaptıkları bir araştırmada, CHAID algoritmasını kullanarak, cidago yüksekliği, vucut uzunluğu ve göğüs çevresi gibi vücut ölçülerinden canlı ağırlık tahmini yapmışlardır. Araştırmada, CHAID regresyon ağacı algoritmasına göre canlı ağırlık üzerinde en etkili değişkenin göğüs çevresi olduğu belirlenmiştir. Sonuç olarak, en ağır koyunların, göğüs çevresi 89 cm’den geniş olan koyunlardan elde edildiği belirlenmiştir.

Eyduran *et al.* (2013b), Türk Saanen keçilerinde laktasyon süt verimi, somatik hücre sayısı ve meme özellikleri arasındaki ilişkiyi ortaya koymak için Exhaustive CHAID algoritması kullanmışlardır. Söz konusu algoritmaya ilişkin regresyon ağacı incelendiğinde, en yüksek süt veriminin, meme açısı 50⁰ dereceden daha geniş keçilerden elde edildiği bildirilmiştir. Meme açısı 50⁰ dereceden daha dar olan keçiler içerisinde laktasyon süresi 191 günden uzun olanların süt veriminin, laktasyon uzunluğu 191 gün ya da 191 günden daha kısa olanlardan (192,25 lt) daha yüksek olduğu belirtilmiştir (Düz. P<0,05). Meme açısı 50⁰ dereceden daha dar olan keçilerin süt verimi, laktasyon süresinden etkilenmektedir (Düz. P<0,05). Ayrıca meme açısı 50⁰ dereceden daha dar ve laktasyon süresi 191 ve 191 günden daha kısa olan keçilerin süt verimi üzerine meme hacminin etkili olduğu belirtilmiştir (Düz.P<0,05).

Eyduran *et al.* (2013c), Mengali koyunlarının doğum ağırlığı üzerinde etkili olan çevre faktörlerini belirlemek amacıyla CHAID algoritmasını kullanmışlardır. Bu çalışmada, birinci dereceden etkili değişkenin doğum tipi olduğu belirtilmiştir. Bu araştırmada, 59 aydan daha yaşlı analardan doğan kuzuların en yüksek doğum ağırlığına sahip olduğu ve ana yaşı faktörünün tekiz erkek kuzuların doğum ağırlığı üzerinde etkili olduğu belirlenmiştir. Buna karşın, tekiz ve dişi kuzuların doğum ağırlığını etkileyen faktörün doğum yılı olduğu bildirilmiştir.

Yılmaz *et al.* (2013) Brown Swiss buzağularında doğum ağırlığını etkileyen faktörleri tespit etmek amacıyla Exhaustive CHAID algoritmasını kullanmışlardır. Araştırmacılar, bağımsız değişkenler arasındaki etkileşimlerin kolayca anlaşılması ve en iyi sonuçları veren bağımsız değişken kombinasyonların belirlenmesi amacıyla bu algoritmayı kullanmanın avantajlı olabileceğini bildirmişlerdir.

Khan *et al.* (2014), Harnai koyunlarından alınan vücut ölçüleri (vücut uzunluğu, cidago yüksekliği, göğüs çevresi, karın çevresi, yüz uzunluğu, kulaklar arası uzunluk, kuyruk genişliği ve uzunluğu) ile canlı ağırlığı tahmin etmek için Exhaustive CHAID algoritmasını özelleştirmişlerdir. Araştırmanın sonucunda, canlı ağırlığı üzerinde birinci derece etkili bağımsız değişkenin yüz uzunluğu olduğu (Düz. P=0,000), ikinci derece etkili değişkenlerin cidago yüksekliği (Düz. P=0,000) ve göğüs çevresi (Düz. P=0,000), üçüncü derece etkili değişkenlerin göğüs çevresi (Düz. P=0,000) ve vücut uzunluğu (Düz. P=0,001) olduğu belirlenmiştir. Böylece, bu algoritma ile oluşturulan regresyon sonuçlarına bakıldığında, canlı ağırlığa ilişkin belirleme katsayısının % 84.4 R² olduğu bildirilmiştir.

Orhan *et al.* (2016) ticari beyaz yumurtacı hibritlerden elde edilen 2049 yumurtadan elde edilen yumurta iç kalite özelliklerinden (kabuk ağırlığı, sarı ve ak ağırlığı) yumurta ağırlığını tahmin etmişlerdir. Araştırmacılar, yumurta ağırlığının tahmin edilmesi bakımından CHAID algoritmasının mükemmel bir uyum (0.9988 R²) gösterdiği sonucuna varmışlardır.

Akşahan (2015) kültür ve melez tosunlarının besi sonu canlı ağırlığını etkileyen vücut ölçülerini belirlemek amacıyla CHAID veri madenciliği algoritmasını esas alan regresyon ağacı methodu kullanmıştır. Besi sonu canlı ağırlığı bakımından gerçek ve tahmin edilen değerler arasındaki Pearson korelasyon katsayısı 0,937 bulunmuştur. Sonuç olarak, CHAID algoritması ile elde edilen sonuçların pratik anlamda faydalı olduğu görülmüştür.

Ali *et al.* (2015) süttten kesim sonrası dönemde Harnai kuzularının çeşitli vücut ölçülerinden canlı ağırlığı tahmin etmek amacıyla CART, CHAID, Exhaustive CHAID ve RBF algoritmalarını karşılaştırmalı olarak incelemişlerdir. Araştırmada, Exhaustive CHAID ile elde edilen regresyon ağacının gerek tahmin performansı ve gerekse biyolojik olarak daha uygun olduğu belirtilmiştir.

Eyduran (2016) küçükbaş hayvanlarda canlı ağırlığın tahmin edilmesinde veri madenciliği algoritmalarının kullanım olanakları konusuna değinmiştir. Araştırmacı, çoklu regresyonda bazı temel varsayımların yerine gelmediği ve çoklu bağlantılılığın

görüldüğü durumlarda CART, CHAID, Exhaustive CHAID, MLP ve MARS gibi algoritmalarının çoklu regresyon analizine alternatif olacağını bildirmiştir.

Eyduran *et al.* (2016) Akkaraman ve İvesi ırkı koyunları üzerinde yürüttükleri bir çalışmada, bazı yaprağı özelliklerinden kirli yaprağı ağırlığını tahmin etmek amacıyla CHAID algoritmasını kullanmışlardır. Araştırmacılar, genetik seleksiyon çalışmalarında verim özelliğini etkileyen değişkenlerin tespit edilmesi bakımından CHAID algoritmasının faydalı olabileceğini ifade etmişlerdir.

Akkuş (2017) koyunlarda canlı ağırlığı tahmin etmek amacıyla MARS veri madenciliği algoritmasını kullanmıştır. Araştırmacı, 26 ve 38 terim içeren MARS modellerinin, canlı ağırlık özelliği bakımından gerçek ve tahmin edilen değerler arasında oldukça yüksek korelasyon ($r=0.980$ ve 0.990) bulmuştur. Sonuç olarak, çalışılan koyunlara ilişkin ırk standartlarının belirlenmesi bakımından MARS algoritmasından elde edilen sonuçların önem arz ettiği ifade edilmiştir.

Duru *et al.* (2017) dört farklı dozda kurutulmuş havuç yaprağı tozu ile beslenen Hy-line Beyaz yumurtacı tavuklar için yumurta iç ve dış özellikleri arasındaki ilişkiyi CHAID algoritması ile incelemişlerdir. Araştırmacılar, CHAID algoritması ile elde edilen regresyon ağacının, yumurta şekil indeksi kapsamında yumurta kartonları ve taşıma için uygun olan yumurtaların hangi en ve yükseklikte olması gerektiği konusunda bazı yararlı ipuçları verdiğini bildirmişlerdir.

Eyduran *et al.* (2017a) Pakistan'da yetiştirilen Beetal keçilerinden ölçülen morfolojik özelliklerden canlı ağırlığı tahmini yapmak amacıyla OLS, CART, CHAID, RBF, MLP1 ve MLP2 algoritmalarını kullanmışlardır. Araştırmacılar, ele alınan algoritmaların canlı ağırlık tahmini yapılmasının ırk standartlarının belirlenmesi, uygun ilaç dozunun ve yem miktarlarının belirlenmesi açısından yararlı olabileceğinden bahsetmişlerdir.

Eyduran *et al.* (2017b) Mengali koçları üzerinde yaptıkları çalışmada, vücut uzunluğu, cidago yüksekliği ve göğüs çevresi özelliklerinden canlı ağırlık özelliğini tahmin etmek amacıyla MARS algoritmasını kullanmışlardır. Araştırmacılar, MARS algoritması ile canlı ağırlığı tahmin etmek için $CA = 34.53158 + 0.57169 * \max(0; GÇ -$

$83.82) -0.54473*\max(0; 83.82-G\check{C}) -0.01059*\max(0; VU-68.58)*\max(0; G\check{C}-83.82)$ şeklinde bir tahmin denklemi geliřtirmişlerdir.

Karabacak *et al.* (2017) Konya ilinde yetiřtirilen Akkaraman (9 baş), Dađlıç (10 baş), Kivircik (10 baş), Merinos (10 baş) ve Karacabey Merinosu (8 baş) koyunlarında sođuk karkas ađırlığı ve canlı ađırlık tahminleri yapmak amacıyla CART, CHAID ve Exhaustive CHAID algoritmalarını kullanmışlardır. Tahmin performansı açısından CHAID ve CART algoritmalarının birbirine benzer sonuçlar ürettiđi bu çalıřmada, CHAID algoritmasının biyolojik anlamda daha uygun ađaç yapısı ürettiđi sonucuna varılmıştır.

Takma ve ark. (2017) yumurta verimini etkileyen özellikleri belirlemek amacıyla CART veri madenciliđi algoritmasını esas alan regresyon ađacı kullanmışlardır. Arařtıřıcılar, karmařık veri setlerinden mümkün olduđunca homojen altgruplar oluřturması nedeniyle, CART algoritmasının kullanılmasının yararlı olacađını belirtmişlerdir.

Aytekin *et al.* (2018) kültür ırkı ve melez olan 103 baş genç bođadan alınan vücut ölçüleri ve besi süresinden yola çıkarak besi sonu canlı ađırlığını tahmin etmek için MARS veri madenciliđi algoritmasını kullanmışlardır. Arařtıřıcılar, besi sonu canlı ađırlığı etkileyen faktörlerin ortaya konulması, söz konusu genotiplere iliřkin standartların belirlenmesi ve en uygun besi süresinin tespit edilmesi açısından MARS algoritması kullanılmasının yararlı olacađını belirtmişlerdir.

Celik *et al.* (2017a) Mengali koçlarından alınan bazı testis ve vücut özelliklerinden canlı ađırlığı tahmin etmek için CART, CHAID, Exhaustive CHAID, MARS, MLP ve RBF algoritmalarını kullanmışlardır. Arařtıřıcılar, tahmin performansı bakımından algoritma sıralamasının $CART > CHAID \approx Exhaustive CHAID > MARS_2 > MARS_1 > RBF > MLP$ şeklinde olduđunu tespit etmişlerdir. Tahmin performansı bakımından etkin algoritmaların kullanılmasının ileride yapılacak ıřlah çalıřmalarında isabetli kararların alınması açısından faydalı olabileceđi ifade edilmiştir. Sonuç olarak bu çalıřmada, canlı ađırlık ile iliřkili morfolojik özelliklerin belirlenmesi gerek dolaylı seleksiyon açısından ve gerekse ırk standartlarının ortaya konulması açısından önem arz ettiđi bildirilmiştir.

Celik *et al.* (2017b) Japon bildircinlarının iç ve dış yumurta kalite özelliklerinden yumurta ağırlığını tahmin etmek amacıyla CART, CHAID ve Exhaustive CHAID algoritmalarını karşılaştırmalı olarak incelemişlerdir. Araştırmacılar, bazı uyum iyiliği testleri (CV%, SS oranı, RAE, RMSE, R^2 ve Düz R^2) bakımından, yumurta ağırlığının tahmin edilmesinde Exhaustive CHAID algoritmasının daha iyi sonuçlar verdiğini bildirmişlerdir.

Celik and Yilmaz (2017) farklı Kangal köpeği varyetelerinden ölçülen farklı vücut ölçülerini kullanarak canlı ağırlık tahmini yapmak amacıyla CART, CHAID ve Exhaustive CHAID veri madenciliği algoritmalarını kullanmışlardır. Araştırmacılar, canlı ağırlığa ilişkin toplam varyasyonu açıklayan en iyi algoritmanın CHAID algoritması olduğunu bildirmişlerdir.

Celik and Yilmaz (2018) Türk Tazı köpeklerinin bazı morfolojik özelliklerinden canlı ağırlık tahmini yapmak amacıyla MARS ve CART algoritmalarını kullanmışlardır. Araştırmacılar, MARS algoritmasının diğerine nazaran daha yüksek tahmin doğruluğuna sahip olduğunu bildirmişlerdir.

Karadas *et al.* (2017a) Akkaraman ırkı koyunların laktasyon süt verimini etkileyen faktörleri belirlemek amacıyla EKK, CART, CHAID, Exhaustive CHAID ve MLP algoritmalarını karşılaştırmalı olarak incelemişlerdir. Araştırmacılar, incelenen algoritmaların karşılaştırılmasında; laktasyon süt verimi bakımından gözlenen ve tahmin edilen değerler arasındaki Pearson korelasyon (r), R^2 , Düz. R^2 , SSoranı, VK(%), HKOK, GNYH, OMYH ve OMS uyum iyiliği ölçütlerini kullanmışlardır. Bu çalışmanın sonucunda, laktasyon süt veriminin tahmin edilmesinde kullanılan algoritmaların tahmin performans sıralaması, CHAID=Exhaustive CHAID >MLP>OLS>CART şeklinde bulunmuştur.

Koc *et al.* (2017) 1884 baş Mengali kuzusunda süttten kesim ağırlığı tahmini yapmak amacıyla; farklı ebeveyn ve yavru düğüm için CART, CHAID ve Exhaustive CHAID veri madenciliği algoritmalarının performanslarını test etmişlerdir. Araştırmacılar, özelleştirilen düğümler için CHAID algoritmasının gerek biyolojik anlamda gerekse yorumlama kolaylığı açısından daha iyi regresyon ağacı yapısı ürettiği sonucuna varmışlardır.

Eyduran and Türkoğlu (2017) Iğdır ilindeki bazı göçmen kuşların morfolojik özellikleri ile canlı ağırlık özelliği arasındaki ilişkiyi tanımlamak için MARS algoritmasını kullanmışlardır. Araştırmacılar, incelenen kuşlara ilişkin tür standartlarının ortaya konulması bakımından MARS algoritması kullanımının faydalı olabileceğini vurgulamışlardır.

Eyduran *et al.* (2017c), MARS tahmin modellerinde standart sapma oranına bağlı olarak GCV'ye ait üst sınırının bulunması konusunda bir yaklaşım ortaya atmışlardır. Araştırmacılar, R Studio paket programının earth paketinde penalty= -1 ve backward budama metodunu kullanmışlardır. Sonuç olarak bu çalışmada, araştırmacıların GCV üst sınır değeri konusunda bilgi sahibi olmasının doğru bir MARS modeli kurmada faydalı olabileceği vurgulanmıştır. Bu kapsamda, earth paketinde penalty = -1 için küçük “nprune” (istenilen terim sayısı) ile büyük “nk” (maksimum terim sayısı) tanımlanması durumunda daha iyi bir MARS çözümü sağlanabileceği bildirilmiştir (Milborrow, 2018).

Eyduran *et al.* (2017d) CART ve MARS algoritmalarını tahmin performansı bakımından mukayese etmişlerdir. Araştırmacılar, üzerinde durulan kantitatif özellik bakımından iki algoritmanın karşılaştırılmasında; gözlenen ve tahmin edilen değerler arasındaki Pearson korelasyon (r), R^2 , Düz. R^2 ve standart sapma oranı uyum iyiliği ölçütlerini kullanmışlardır. Bu çalışmada, optimizasyon kapsamında MARS algoritmasının yanıt yüzeyi metoduna alternatif olabileceği bildirilmiştir. Değişkenlerin dağılımları ile ilgili varsayımların sağlanmaması durumunda veri madenciliği algoritmalarının EKK'dan daha informatif olabileceği vurgulanmıştır.

Oguntunji (2017) Nijerya Muscovy ördeklerinden alınan vücut ölçülerinden canlı ağırlık özelliğini tahmin etmek amacıyla CART veri madenciliği algoritmasına dayalı regresyon ağacı metodunu kullanmıştır. Araştırmacı, CART ile oluşturulan regresyon ağacı analizi ile canlı ağırlığa ait toplam farklılığın % 91'inin (R^2) açıklandığını bildirmiştir.

Tahtalı ve Berberoğlu (2017) Polatlı Tarım işletmesinde yetiştirilen 894 Holstein-Fresian buzağı üzerinde yürüttükleri bir çalışmada doğum ağırlığını etkileyen faktörleri belirlemek amacıyla CHAID algoritması kullanmışlardır.

Balta and Topal (2018), Artvin ilinde yetiştirilen Hemşin kuzularında doğum ağırlığını etkileyen faktörleri belirlemek amacıyla CART algoritmasına dayalı regresyon ağacı metodunu kullanmışlardır. Araştırmacılar, risk değeri (0.456) ve R^2 (0.862) uyum iyiliği ölçütlerine göre CART algoritmasının tahmin performansının iyi olduğunu belirtmişlerdir.

Şengül ve ark. (2018), bildircınlar üzerinde yürüttükleri çalışmada, göğüs etinde parlaklığı etkileyen faktörleri belirlemek amacıyla MARS algoritması kullanarak interaksiyon terimleri içeren bir tahmin denklemi elde etmişlerdir. Araştırmacılar, belirleme ve düzeltilmiş belirleme katsayılarını 0.93 ve 0.89 olarak bulmuşlardır.

Ertürk (2018) Doğu Anadolu'da yetiştirilen yerli sığırlarda besi performansını etkileyen faktörleri belirlemek amacıyla interaksiyon terimleri içeren MARS algoritmasını kullanmıştır. Araştırmacı, MARS algoritması sonuçlarına göre sosyal, ekonomik ve biyolojik faktörlerin besiciliği etkileyen unsurlar olduğunu bildirmiştir.

Ertürk *et al.* (2018) Doğu Anadolu'da yetiştirilen melez sığırlarda besi performansını etkileyen faktörleri tespit etmek amacıyla interaksiyon terimleri içeren MARS algoritmasını kullanmışlardır. Araştırmacılar, besi sonu canlı ağırlığı bakımından gerçek değerler ve MARS ile tahmin edilen değerler arasında çok yüksek korelasyon ($r \sim 1.000$) olduğunu bildirmişlerdir. MARS algoritması sonucunda, sosyal, ekonomik ve biyolojik faktörlerin besiciliği etkileyen unsurlar olduğu belirtilmiştir.

Gevrekçi and Takma (2018) yumurtacı tavuklarda yumurta üretimini etkileyen faktörleri belirlemek amacıyla, CART ve CHAID veri madenciliği algoritmalarını karşılaştırmalı olarak incelemişlerdir. Araştırmacılar, tahmin doğruluğu bakımından (daha düşük CV, GNYH, OMS ve HKOK değeri alan) CHAID algoritmasının CART algoritmasına nazaran daha iyi sonuçlar verdiğini beyan etmişlerdir.

Kaygısız et al. (2018) yarış atlarında yarış süresini etkileyen faktörleri belirlemek amacıyla MARS veri madenciliği algoritmasını kullanmışlardır. Araştırmacılar, aşırı uyum iyiliğini önlemek amacıyla R programının earth paketinde $penalty=2$, $degree=3$, $ifold=10$ ve $nk=300$ tanımlamalarını yapmışlar, belirleme ve düzeltilmiş belirleme katsayıları 0.972 olarak bulmuşlardır. Araştırma sonucunda, "Yarış süresi="

143.63+5.16*kumpist – 3.44*grupyarıřtđpi + 0.08* max(0, kořumesafesi–2000)”
MARS tahmin modeli elde edilmiřtir.

Aksoy *et al.* (2018a) Doęu Anadolu’da yetiřtirilen kiltür ırkı sđęırlarda besi performansını etkileyen faktörleri belirlemek için CHAID (r=0.819) ve MARS (r=0.992) algoritmasını kullanmıřlardır. Arařtırcılar, bu çalıřmada, tahmin performansı bakımından MARS algoritmasının CHAID algoritmasına nazaran daha iyi olduęunu belirtmiřlerdir. Çalıřmada besi bařı canlı aęırlıęının besi sonu canlı aęırlıęı üzerinde daha belirleyici olduęu bildirilmiřtir. Bu çalıřmadan elde edilen sonuçlara göre besi performansı aısından sosyal, ekonomik ve biyolojik faktörlerin birlikte incelenmesi gerektięi sonucuna varılmıřtır.

Aksoy *et al.* (2018b) Aęrı, Kars ve Erzurum illerindeki arıcılık iřletmelerinde kovan bařına verimini etkileyen faktörleri belirlemek amacıyla CART, CHAID ve MARS veri madencilięi algoritmalarını karřılařtırmalı olarak incelemiřlerdir. Bu çalıřmada, elde edilen performans sıralaması MARS (r=0.920) > CART (r=0.619) > OLS (r=0.286) řeklinde olmasına karřın, CHAID algoritmasından herhangi bir sonuç elde edilememiřtir. Sonuç olarak, bu çalıřmada, kovan bařına verim aısından sosyal, ekonomik ve biyolojik faktörlerin birlikte incelenmesinin gerekli olduęu vurgulanmıřtır.

Aksoy *et al.* (2018c) yarı entansif kořullarda yetiřtirilen Morkaraman ve Romanov x Morkaraman F₁ melezi kuzularından elde edilen mera gelirini tahmin etmek amacıyla MARS veri madencilięi algoritmasını kullanmıřlardır. Arařtırcılar, en dűřük GCV deęerini veren ve interaksiyon terimi iermeyen MARS modelinin tahmin doęruluęunun % 100’e yakın olduęunu bildirmiřlerdir. Sonuç olarak bu çalıřmada, kuzu üretkenlięinin geliřtirilmesi kapsamında MARS algoritması kullanımının, mera gelirini etkileyen faktörlere iliřkin optimal noktaların belirlenmesi konusunda yararlı olabileceęi vurgulanmıřtır.

Celik *et al.* (2018) Daera Din Panah (DDP) keilerinde canlı aęırlık-yař iliřkisini modellemek amacıyla Morgan-Mercer-Flodin (MMF), Logistic, Von-Bertalanffy ve Janosheck modelleri ile birlikte MARS algoritmasını kullanmıřlardır. Bu çalıřmada, MARS algoritmasından Canlı aęırlık=11.26833 + 0.07073*max(0; zaman - 30) tahmin denklemi elde edilmiřtir. Sonuç olarak, MARS algoritmasının (0.970 R² ve

0.960 Düz.R²) söz konusu ilişkinin tanımlanmasında alternatif olabileceği sonucuna varılmıştır.

Olfaz *et al.* (2019) Karayaka kuzularında süttten kesim ağırlığını etkileyen faktörleri belirlemek amacıyla CART ve CHAID veri madenciliği algoritmasını kullanmışlardır. Algoritmaların tahmin performansı açısından (0.88 R²) eş değer olmasına karşın, biyolojik yorumlama bakımından CHAID algoritması ile üretilen regresyon ağacının daha uygun olduğu belirlenmiştir.

Zaborski *et al.* (2018) Balochistan (Pakistan)'da yetiştirilen yerli Harnai kuzularında doğumdan 30. gün, 60. gün, 90. gün ve 180. güne kadar olan periyotlardaki günlük canlı ağırlık kazancını tahmin etmek amacıyla MARS veri madenciliği algoritmasını kullanmışlardır. Araştırmacılar, 7863 kuzulama kaydı bulunan veri setini %75 eğitim seti ve %25 test seti olmak üzere iki alt sete bölmüşlerdir. Araştırmalar sonucunda ikinci derece interaksiyon terimi içeren MARS modellerinin tahminleme performanslarının (r=0.21-0.30) düşük olduğu tespit edilmiştir.

3. MATERYAL ve METOT

3.1. Materyal

Tahmin doğruluğu bakımından CART, CHAID Exhaustive CHAID ve MARS algoritmalarının performanslarını karşılaştırmak amacıyla Pakistan’da yetiştirilen Teddy ırkı keçiler kullanılmıştır. Bu ırka ait canlı ağırlığı (kg) tahmin etmek amacıyla kullanılan sürekli bağımsız değişkenlere (cm) ilişkin tanıtıcı istatistikler Çizelge 3.1’de verilmiştir.

Çizelge 3.1. Sürekli bağımsız değişkenlerin tanıtıcı istatistikleri

Cinsiyet	N	%
Dişi (F)	150	91,5
Erkek (M)	14	8,5
Sürekli Değişkenler	Ortalama	Standart Hata
Canlı Ağırlık (LBW)	24.12	0.48
Sağrı uzunluğu (RUMPLENGTH)	13.50	0.10
Kuyruk Uzunluğu (TAILLENGTH)	15.26	0.17
İncik Çevresi (SHANKCIRCUMFERENCE)	8.71	0.06
Sağrı Yüksekliği (RUMPHEIGHT)	63.90	0.56
Cidago Yüksekliği (WITHERHEIGHT)	68.00	0.55
Baş Uzunluğu (HEADLENGTH)	18.70	0.11
Ağız Genişliği (MOUTHWIDTH)	12.37	0.06
Boynuz Uzunluğu (HORNLENGTH)	15.54	0.37
Kulak Uzunluğu (EARLENGTH)	9.24	0.22
Boyun Uzunluğu (NECKLENGTH)	20.54	0.19
Boyun Çevresi (NECKGIRTH)	30.67	0.30
Vücut Uzunluğu (BODYLENGTH)	60.30	0.39
Göğüs Çevresi (HEARTGIRTH)	69.70	0.44
Karın Çevresi (BELLYGIRTH)	78.50	0.60
Vücut Derinliği (BODYDEPTHHEART)	26.70	0.10
Ön Sağrı Genişliği (RUMPWIDTHFRONT)	13.10	0.09

3.2. Metot

Bu çalışma kapsamında, üzerinde durulan verim (kantitatif) değişkeni üzerinde etkili olan değişkenlerin belirlenmesi kapsamında CART, CHAID ve Exhaustive CHAID algoritmaları kullanılarak karar ağacı yani regresyon ağacı diyagramlarının oluşturulması planlanmıştır. Aynı kapsamda, MARS veri madenciliği algoritması da kullanılmış ve ağaca dayalı algoritmalarla karşılaştırmalı olarak incelenmiştir.

CART, CHAID ve Exhaustive CHAID gibi veri madenciliği algoritmaları ağaç yapısına dayalı algoritmalar olmakla birlikte nominal, ordinal ve sürekli yapıdaki değişkenlerin etkin analizinde rahatlıkla kullanılabilir. Breiman ve ark. (1984) tarafından geliştirilen CART veri madenciliği algoritması, veri setini (learning sample) oluşturan bütün bireyleri içine alan kök düğümden başlayarak, yinelemeli olarak homojen düğümler elde edilene kadar her aşamada bir düğümden sadece iki yavru düğüm şeklinde bölünen ikili karar (Binary node splitting) ağacı algoritmasıdır. Non-parametrik bir yaklaşım olan CART, analiz edilecek veri setinin dağılımı ile ilgili herhangi bir varsayım gerektirmez ve istatistiksel anlamda uyum iyiliğini geliştiren bağımsız değişkenler ile karar ağacı elde edilmesini sağlar (Nisbet ve ark., 2009). CART için her bölünme sadece bir bağımsız değişkenin değerine bağlıdır. I adet kategorisi olan sıralı değişken bakımından $2^{I-1}-1$ adet bölünme mümkündür. Ayrıca, CART algoritması ile elde edilecek bir karar ağacında yorumlama kolaylığının sağlanması açısından tahmin doğruluğu üzerinde olumsuz etki yapabilecek gereksiz dalların ya da çocuk düğümlerin ağaç yapısından kaldırılması için IBM SPSS programında budama (pruning) işlemi aktif hale getirilmelidir. Ancak, IBM SPSS 23 programında cross-validation aktive edildiği durumlarda pruning işleminin uygulanmasına gerek yoktur.

CART algoritmasından elde edilen bir regresyon ağacında aynı değişken bir dalda birkaç kez bölünmeye (surrogate splits) yol açabilmektedir. Diğer bir ifadeyle, CART algoritmasının aşırı dallanma potansiyeline sahip bir algoritma olduğu bilindiğinden dolayı elde edilen regresyon ağacının yorumlanması zor olabilmektedir. Bu gibi durumlardan kaçınmak amacıyla değerlendirilen veri setini öğrenme (Learning, Training) ve test (testing, validation) setlerine ayrılması, ebeveyn-yavru düğümlerde bulunması gereken birey sayısının ve ağaç derinliği (tree depth) sayısının hassas bir

şekilde ayarlanması gerekmektedir. Genelde öğrenme ve test setleri için 80:20 oranının kullanımı tavsiye edilir. Ancak, uyum iyiliğini geliştirmek amacıyla araştırmacılar tarafından farklı oranlarda özelleştirilebilir. Optimum bir ağaç yapısı için bu iki set için tahmin edilen uyum iyiliği ölçütlerinin birbirine yakın olması önerilmektedir. Diğer bir seçenek olarak iki set oluşturmak yerine çapraz geçerlik (validation) 10 alınıp mevcut veri seti için ortalama bir hata değeri hesaplanabilir. Bir bireye ait bağımlı değişkene ait değer ya da tüm bağımsız değişkenlere ait değerler kayıpsa, CART algoritmasında bu birey analize dâhil edilmez.

Kass (1980) tarafından geliştirilen CHAID algoritması bir düğümden iki ya da ikiden fazla yavru düğüm olacak şekilde bölünmeye imkan vermektedir. CHAID ve Exhaustive CHAID algoritmaları, sürekli bağımsız değişkenleri sıralı (ordinal) değişkenlere dönüştürür ve her aşamada bağımlı değişken ile interaksiyona girecek bağımsız değişkenleri regresyon ağacı yapısına dahil etmektedir. CHAID veri madenciliği algoritması, veri setindeki tüm bireylerin bulunduğu kök düğümden başlayarak, yinelemeli olarak homojen düğümler elde edilene kadar her aşamada bir düğümden iki ya da ikiden fazla sayıda yavru düğüm olacak şekilde bölünen çoklu bölünme (multiway node splitting) ağacı algoritmasıdır. CHAID algoritmaları birleştirme, ayırma ve durdurma aşamaları ile regresyon ağacı oluşturulmaktadır. Exhaustive CHAID algoritması, CHAID algoritmasına benzer olup, sadece birleştirme aşamaları bakımından farklıdır.

IBM SPSS 23 programında CHAID algoritması ile oluşturulan regresyon ağacında oluşabilecek gereksiz dallanmaları engellemek için pruning işlemi otomatik yapılmaktadır. Yapılan araştırmalarda, biyolojik verilerde daha uygun regresyon ağaç yapısı üretmesi hasebiyle CHAID algoritmalarının kullanımı önerilmektedir (Ali ve ark. 2015).

Uyum iyiliğinin geliştirilmesi açısından gerek CART gerekse CHAID algoritmaları için analiz edilmesi planlanan veri setinin öğrenme (Learning, Training) ve test (testing, validation) setlerine ayrılması, ebeveyn-yavru düğümden bulunması gereken birey sayısının ve ağaç derinliği (tree depth) sayısının dikkatli bir biçimde ayarlanması gerekmektedir.

Ünlü fizikçi Jerome Friedman (1991) tarafından geliştirilen MARS algoritması, bağımlı ve bağımsız değişken setleri arasındaki ilişkinin en etkin biçimde tanımlanmasını sağlayan bir non-parametrik regresyon yöntemidir. MARS algoritması bağımlı ve bağımsız değişkenler arasındaki fonksiyonel ilişkiler ile ilgili varsayım gerektirmemektedir. MARS algoritması için tüm değişkenlerin dağılımı konusunda herhangi bir varsayıma gerek yoktur. Eğitim veri setleri (training data sets), değişik eğimleri olan parçalı doğrusal segmentlere yani splaynlara ayrılmaktadır. Splaynlar birbirine bağlıdır ve parçalı eğriler olarak da ifade edilen temel fonksiyonlar (basis functions), doğrusal ve doğrusal olmayan etkilerin daha doğru tanımlanmasına olanak sağlar. Splaynlar arasındaki bağlantı noktaları, düğümler (knots) olarak adlandırılır. Aday düğümler, her bir bağımsız değişkenin değişim sınırları içerisinde bulunur. MARS algoritması, adimsal süreç (stepwise procedure) kapsamında mümkün olan tüm düğümleri ve interaksiyon etkilerini hesaba katarak temel fonksiyonları türetmektedir. Temel fonksiyon çiftlerini tanımlamak amacıyla değişken ekleme (forward procedure) prosedürü, her bir bağımsız değişkenin değişim sınırları içerisinde rasgele bir yerde olacak şekilde aday düğümlerin belirlenmesini sağlar. Her bir aşamada, MARS algoritması tarafından oluşturulan model, en düşük hata varyansının elde edilmesini sağlayan düğümleri ve temel fonksiyon çiftlerini özelleştirir. Karmaşık modele ulaşıldığında, temel fonksiyonların dahil edilmesi devam eder. MARS ile oluşturulan modelde, tahmin performansını düşüren gereksiz fonksiyonlar geri eleme prosedürü (backward procedure) ile kaldırılır (Zhang and Goh, 2016).

MARS için tahmin modeli aşağıdaki gibi yazılabilir:

$$\hat{y} = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} h_{km}(X_{v(k,m)}) \quad (3.1.)$$

Burada;

\hat{y} : Bağımlı değişkenin tahmin değeri, β_0 sabit, $h_{km}(X_{v(k,m)})$ temel fonksiyon,

k. çarpanının m. component bakımından bağımsız değişkene ait indeks,

K_m is interaksiyon derecesini sınırlayan parameter değerini ifade etmektedir.

Budama algoritması, genelleştirilmiş çapraz geçerlilik (generalized cross-validation error, GCV) tekniği ile yapılır (Kornacki and Ćwik 2005). GCV hem artıkların hatasını hem de model kompleksliğini dikkate almaktadır.

$$GCV(\lambda) = \frac{\sum_{i=1}^n (y_i - y_{ip})^2}{\left[1 - \frac{M(\lambda)}{n}\right]^2} \quad (3.2.)$$

Burada:

n : Eğitim setindeki hayvan sayısı,

y_i : i . hayvana ait bağımlı değişken değeri,

y_{ip} : i . hayvana ait bağımlı değişken tahmin değeri,

$M(\lambda)$: λ terimlerini içeren modelin karmaşıklığına ilişkin ceza fonksiyonunu göstermektedir.

Araştırmaya konu olan algoritmaların performanslarını karşılaştırmak için kullanılacak model uyum iyiliği ölçütleri aşağıda verilmiştir (Eyduran ve ark., 2017):

1. Bağımlı değişken bakımından gerçek değerler ve tahmin edilen değerler arasındaki Pearson korelasyon katsayısı (r), (3.3)
2. Akaike bilgi ölçütü (Akaike information criterion, AIC)

$$AIC = n \ln \left[\frac{1}{n} \sum_{i=1}^n (y_i - y_{ip})^2 \right] + 2k, \text{ eğer } n/k > 40 \quad (3.4.)$$

$$\text{ya da : } AIC_c = n \ln \left[\frac{1}{n} \sum_{i=1}^n (y_i - y_{ip})^2 \right] + 2k + \frac{2k(k+1)}{n-k-1}, \text{ aksi halde} \quad (3.5.)$$

3. Hata kareler ortalamasının karekökü (Root-mean-square error, RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{ip})^2} \quad (3.6.)$$

4. Ortalama hata (Mean error, ME) :

$$ME = \frac{1}{n} \sum_{i=1}^n (y_i - y_{ip}) \quad (3.7.)$$

5. Ortalama mutlak sapma (Mean absolute deviation, MAD):

$$MAD = \frac{1}{n} \sum_{i=1}^n |y_i - y_{ip}| \quad (3.8.)$$

6. Standart sapma oranı (Standard deviation ratio, SD_{ratio}):

$$SD_{ratio} = \frac{s_m}{s_d} \quad (3.9.)$$

7. Global nispi yaklaşık hata (Global relative approximation error, RAE):

$$RAE = \sqrt{\frac{\sum_{i=1}^n (y_i - y_{ip})^2}{\sum_{i=1}^n y_i^2}} \quad (3.10.)$$

8. Ortalama mutlak yüzde hata (Mean absolute percentage error, MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_{ip}}{y_i} \right| \cdot 100 \quad (3.11.)$$

Burada:

n : Veri setinde bulunan hayvan sayısı,

k : Model parametrelerinin sayısı,

y_i : i . hayvana ait bağımlı deęişken deęeri,

y_{ip} : i . hayvana ait bağımlı deęişken tahmin deęeri,

s_m : Model hata terimlerinin standart sapması ve

s_d : Bağımlı deęişkenin standart sapmasını ifade etmektedir.

CART, CHAID ve Exhaustive CHAID algoritmalarına ilişkin analizler IBM SPSS 23 programı (Analysis > Classify > Tree) ile yapılmıştır. Her üç algoritma için en yüksek tahmin performansına ulaşmak amacıyla ebeveyn çocuk düğümdeki minimum hayvan sayıları 4:2 olarak ayarlanmıştır. Ağaç derinlikleri CART için 5, CHAID algoritmaları için 3 olarak alınmıştır. Her üç algoritma için elde edilen bağımlı deęişken tahmin deęerleri (predicted values) ve gerçek bağımlı deęişken deęerleri (observed values) bir not defteri dosyasında kaydedildikten sonra bir R script dosyası oluşturularak uyum iyilięi ölçütleri R STUDIO programında hesaplanmıştır. Bu çalışmada yazılan R script dosyası tezin ek kısmında sunulmuştur (EK1). MARS algoritmasına ilişkin analizler R STUDIO programının earth paketi (Milborrow, 2018) kullanılarak yapılmıştır (R Core Team, 2014). Hem MARS analizi hemde MARS 'a ait uyum iyilięi ölçütlerini hesaplamak amacıyla R script dosyası oluşturulmuştur (EK2). R STUDIO programında MARS algoritması için,

```
marsresult=earth(LBW~., data=mydata1, penalty=-1, degree=4, pmethod="backward",  
nk=1000, nfold=10, nprune=45, keepxy=T)
```

tanımlaması yapılmıştır. Etkili deęişkenlerden oluşturulan terimlere ait katsayıların önem testlerini yapmak için,

```
bx<-model.matrix(marsresult)  
a.lm<-lm(mydata1$LBW~bx[,-1])  
summary(a.lm)  
tanımlamaları yapılmıştır.
```

4. BULGULAR ve TARTIŞMA

4.1. Algoritmaların Karşılaştırılması

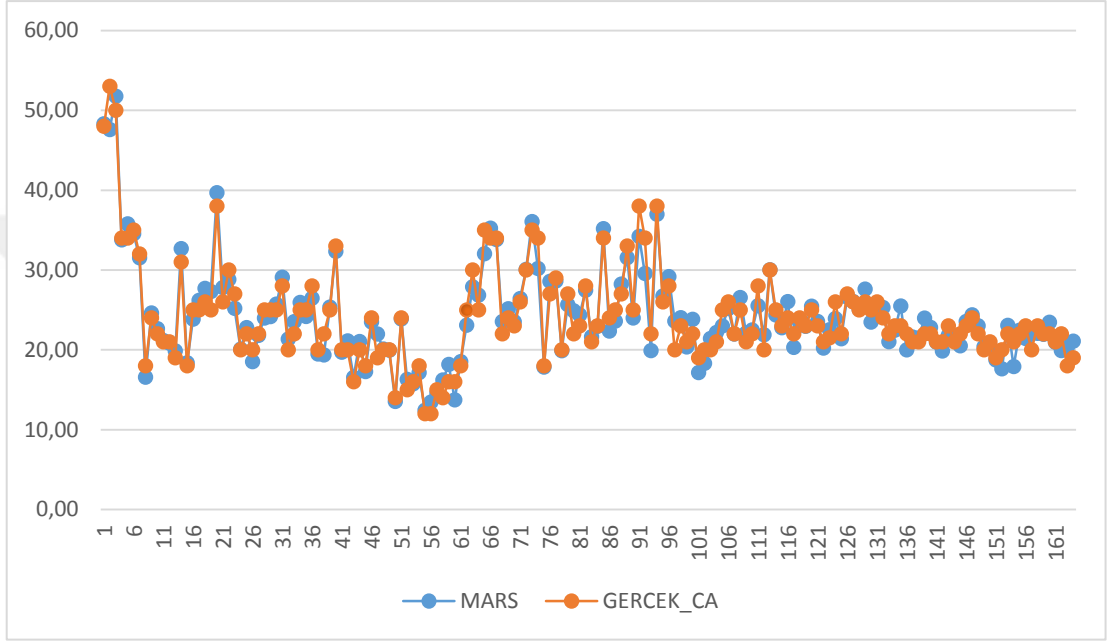
Çalışmada incelenen veri madenciliği algoritmalarına ilişkin hesaplanan uyum iyiliği ölçütleri Çizelge 4.1’de verilmiştir. Algoritmalara ait uyum iyiliği ölçütleri incelendiğinde bağımlı değişkenin (canlı ağırlık) tahmin edilmesi bakımından MARS algoritmasının en iyi algoritma olduğu tespit edilmiştir (Çizelge 4.1). MARS algoritmasından sonra tahmin performansı en iyi olan veri madenciliği algoritmasının Exhaustive CHAID olduğu belirlenmiştir. Bağımlı değişken bakımından gerçek ve tahmin değerler arasındaki Pearson korelasyon katsayılarına bakıldığında da, MARS algoritmasının daha üstün olduğu dikkat çekmektedir ($P < 0.01$).

Çizelge 4.1. Algoritmalara ilişkin uyum iyiliği ölçütleri

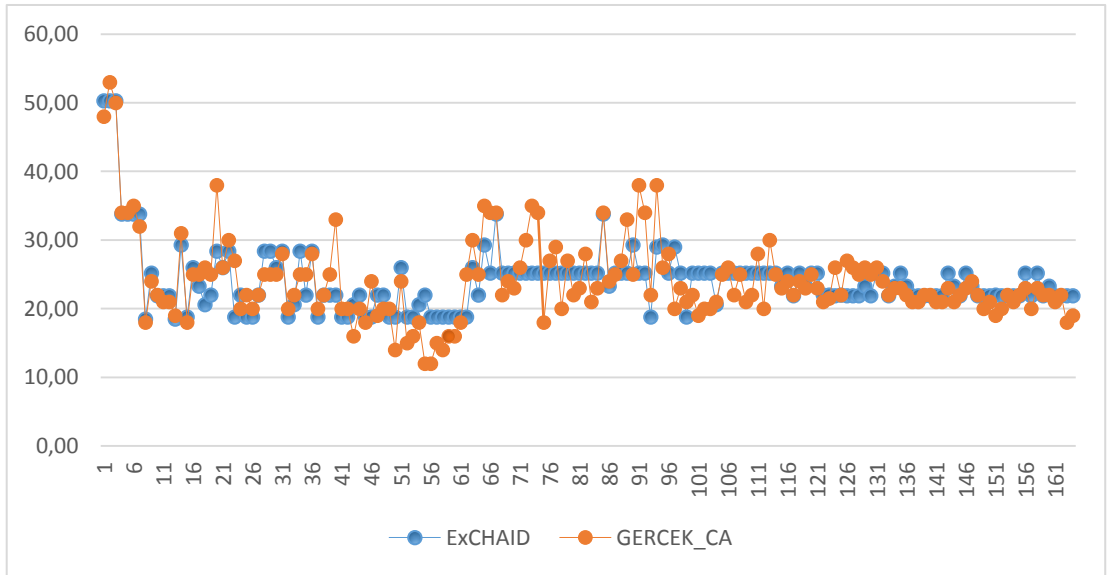
Uyum iyiliği Ölçütleri	MARS	CHAID	CART	ExCHAID
AIC	219.50	456.66	453.65	436.87
Düz. AIC (AICc)	254.58	457.38	453.80	437.40
RMSE	1.484	3.856	3.915	3.652
OH (ME)	0.000	-0.001	-0.002	-0.005
MAD	1.173	2.765	2.857	2.629
SS oranı (SD ratio)	0.242	0.629	0.638	0.595
NYH (RAE)	0.06	0.155	0.157	0.147
OMYH(MAPE)	4.984	12.110	12.580	11.381
VK (CV%)	6.170	16.030	16.280	15.180
R ²	0.941	0.605	0.593	0.646
Düz. R ²	0.918	0.587	0.585	0.632
R	0.970 ^a	0.778 ^b	0.770 ^b	0.803 ^b

Tahmin performansları bakımından incelenen algoritmaların gerçek ve tahmin edilen değerler arasındaki uyumu gösteren grafikler sırasıyla Şekil 4.1, 4.2, 4.3 ve 4.4’de gösterilmiştir. Bağımlı değişken bakımından gerçek ve tahmin edilen değerler arasındaki uyumluluk değerlendirildiğinde, en iyi uyumun MARS algoritması ile sağlandığı belirlenmiştir (Şekil 4.1, 4.2, 4.3 ve 4.4). Ancak, ağaç yapısına dayalı

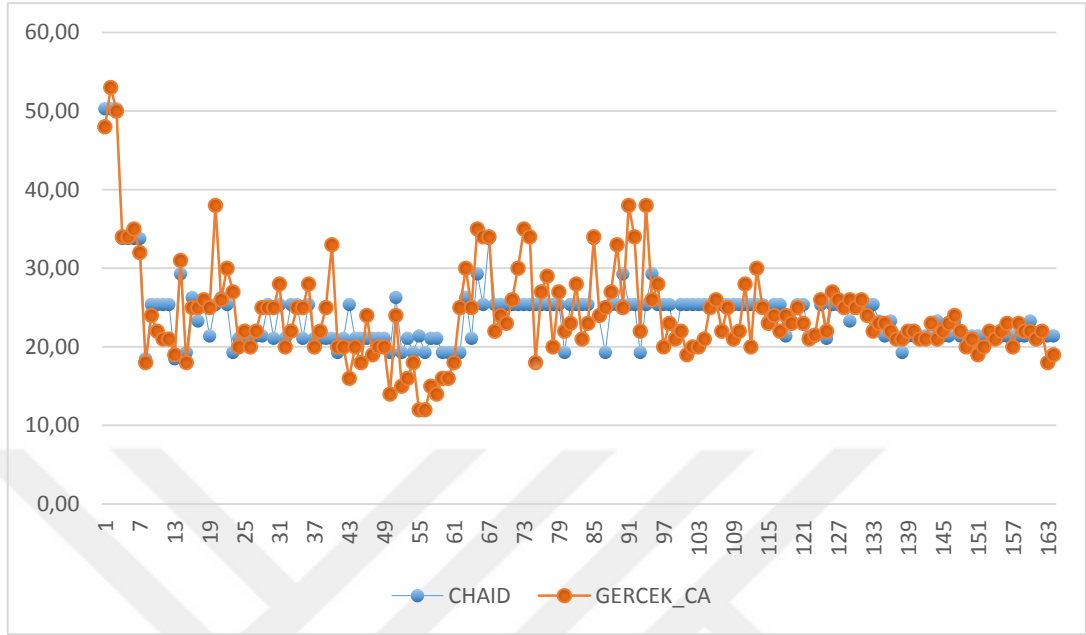
algoritmalar için uyum grafikleri incelendiğinde, uyum sıralamasının Exhaustive CHAID > CHAID > CART olduğu görülmüştür (Şekil 4.2, 4.3 ve 4.4). Diğer bir ifadeyle, CART algoritması için hayvanların çoğunda elde edilen tahmin değerlerinin gerçek değerlerini yansıtmadığı ifade edilebilir.



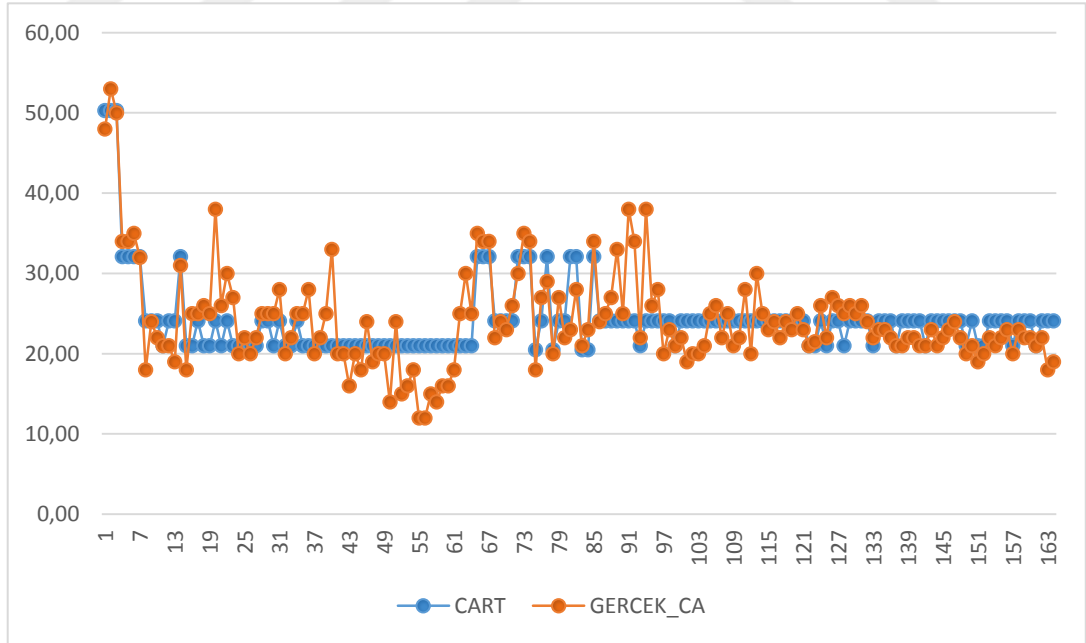
Şekil 4.1. MARS algoritması için uyum grafiği



Şekil 4.2. Exhaustive CHAID algoritması için uyum grafiği



Şekil 4.3. CHAID algoritması için uyum grafiği



Şekil 4.4. CART algoritması için uyum grafiği

4.2. Exhaustive CHAID Algoritmasına İlişkin Sonuçlar

Exhaustive CHAID algoritmasıyla oluşturulan regresyon ağaç diyagramı Şekil 4.5'de gösterilmiştir. Ağaç yapısı incelendiğinde, ön sağrı genişliği (RUMPWIDTHFRONT), sağrı yüksekliği (RUMPHEIGHT), cidago yüksekliği (WITHERHEIGHT), kuyruk uzunluğu (TAILLENGTH), arka sağrı genişliği (RUMPWIDTHBACK) ve cinsiyet (SEX) bağımsız değişkenlerinin canlı ağırlık üzerinde etkili olduğu belirlenmiştir (Düz. $P < 0.05$). Regresyon ağaç diyagramının en üstünde bulunan kök düğüm ön sağrı genişliği bakımından üç alt gruba (Düğüm 1, 2 ve 3) ayrılmıştır (Düz. $P = 0,000$, $F = 28,750$, $sd1 = 2$ ve $sd2 = 161$).

Düğüm 1, ön sağrı genişliği ≤ 14 cm olan hayvanların oluşturduğu alt grubu ifade etmektedir. Düğüm 1'de bulunan hayvanlar toplam hayvan sayısının %84.8'ini oluşturmuşlardır. Düğüm 2, ön sağrı genişliği $14 \text{ cm} < \text{ön sağrı genişliği} \leq 14,50$ cm olan hayvanların oluşturduğu alt grubu ifade etmektedir. Düğüm 3, ön sağrı genişliği $> 14,50$ cm olan hayvanların oluşturduğu alt grubu ifade etmektedir.

Düğüm 4, ön sağrı yüksekliği ≤ 14 cm ve sağrı yüksekliği $\leq 59,50$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 4 kuyruk uzunluğu özelliği bakımından 3 alt gruba (Düğüm 11, 12 ve 13) ayrılmaktadır (Düz. $P = 0,039$, $F = 8,109$, $sd1 = 2$ ve $sd2 = 41$). Düğüm 11, ön sağrı yüksekliği ≤ 14 cm, sağrı yüksekliği $\leq 59,50$ cm ve kuyruk uzunluğu $\leq 12,5$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 12, ön sağrı yüksekliği ≤ 14 cm, sağrı yüksekliği $\leq 59,50$ cm ve kuyruk uzunluğu $12,50 < \text{kuyruk uzunluğu} \leq 15,5$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 13, ön sağrı yüksekliği ≤ 14 cm, sağrı yüksekliği $\leq 59,50$ cm ve kuyruk uzunluğu $> 15,5$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir.

Düğüm 5, ön sağrı yüksekliği ≤ 14 cm, sağrı yüksekliği $59,50 < \text{sağrı yüksekliği} \leq 62,50$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir (23.114 kg). Düğüm 5, arka sağrı genişliği bakımından dört alt gruba (Düğüm 14, 15, 16 ve 17) ayrılmıştır (Düz. $P = 0,002$, $F = 9,692$, $sd1 = 3$ ve $sd2 = 40$). Düğüm 14, ön sağrı yüksekliği ≤ 14 cm, $59,50 < \text{sağrı yüksekliği} \leq 62,50$ cm ve arka sağrı genişliği $\leq 3,50$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 15, ön sağrı yüksekliği ≤ 14 cm, sağrı yüksekliği $59,50 < \text{sağrı yüksekliği} \leq 62,50$ cm ve $3,50 < \text{arka sağrı}$

geniřliđi $\leq 6,50$ cm olan hayvanların oluřturduđu alt grubu temsil etmektedir. Dűđüm 16, ön sađrı yűkseklđi ≤ 14 cm, sađrı yűkseklđi $59,50 < \text{sađrı yűkseklđi} \leq 62,50$ cm ve $6,50 < \text{arka sađrı geniřliđi} \leq 10$ cm olan hayvanların oluřturduđu alt grubu temsil etmektedir. Dűđüm 17, ön sađrı yűkseklđi ≤ 14 cm, sađrı yűkseklđi $59,50 < \text{sađrı yűkseklđi} \leq 62,50$ cm ve arka sađrı geniřliđi > 10 cm olan hayvanların oluřturduđu alt grubu temsil etmektedir.

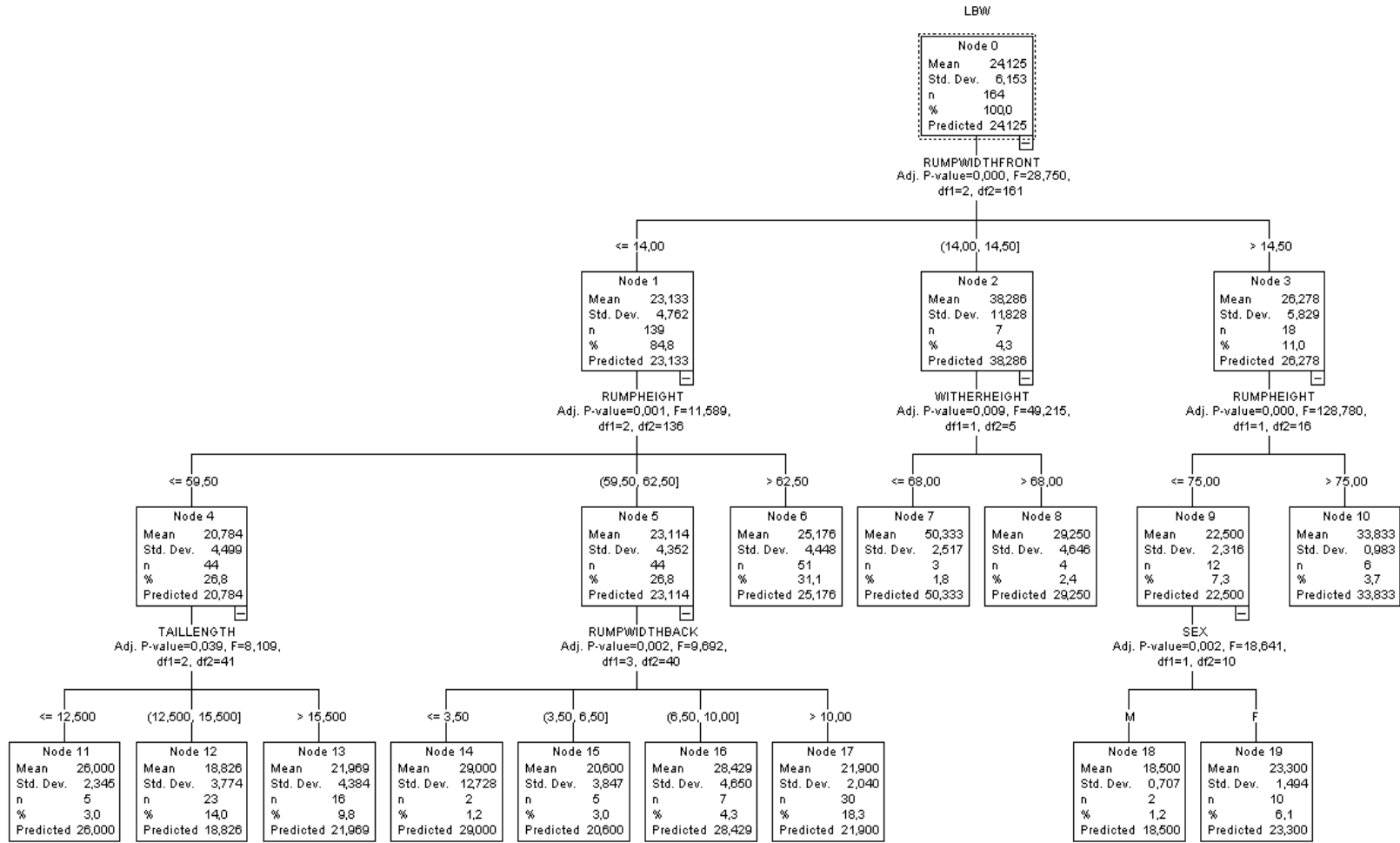
Dűđüm 6, ön sađrı geniřliđi ≤ 14 cm ve sađrı yűkseklđi $> 62,50$ cm olan hayvanların oluřturduđu alt grubu temsil etmektedir (25.176 kg). Dűđüm 6' y' oluřturan hayvanlar yeterince homojen bir yapıya ulařtıđı için herhangi bir bűlűnmeye uđramamıřtır.

Dűđüm 7, $14 < \text{ön sađrı yűkseklđi} \leq 14,50$ cm ve cidago yűkseklđi ≤ 68 olan hayvanların oluřturduđu alt grubu temsil etmektedir. Dűđüm 7' y' oluřturan hayvanlar yeterince homejen bir yapıya ulařtıđı için herhangi bir bűlűnmeye uđramamıřtır.

Dűđüm 8, ön sađrı geniřliđi $14 < \text{ön sađrı geniřliđi} \leq 14,50$ cm ve cidago yűkseklđi > 68 olan hayvanların oluřturduđu alt grubu temsil etmektedir. Dűđüm 8'i oluřturan hayvanlar yeterince homejen bir yapıya ulařtıđı için herhangi bir bűlűnmeye uđramamıřtır.

Dűđüm 9, ön sađrı geniřliđi $>14,50$ cm ve sađrı yűkseklđi ≤ 75 cm olan hayvanların oluřturduđu alt grubu ifade etmektedir. Dűđüm 9, cinsiyet bakımından iki alt gruba (Dűđüm 18 ve 19) ayrılmıřtır (Dűz. $P=0,002$, $F=18,641$, $sd1=1$ ve $sd2=10$). Dűđüm 18, ön sađrı geniřliđi $>14,50$ cm, sađrı yűkseklđi ≤ 75 ve cinsiyeti erkek olan havyanların oluřturduđu alt grubu temsil etmektedir. Dűđüm 19, ön sađrı geniřliđi $>14,50$ cm, sađrı yűkseklđi ≤ 75 ve cinsiyeti diři olan havyanların oluřturduđu alt grubu temsil etmektedir.

Dűđüm 10, ön sađrı geniřliđi $>14,50$ cm ve sađrı yűkseklđi > 75 cm olan hayvanların oluřturduđu alt grubu temsil etmektedir. Dűđüm 10'u oluřturan hayvanlar yeterince homojen bir yapıya ulařtıđı için herhangi bir bűlűnmeye uđramamıřtır.



Şekil 4.5. Exhaustive CHAID algoritması ile oluşturulan ağaç diyagramı

4.3. CHAID Algoritmasına İlişkin Sonuçlar

CHAID algoritmasıyla oluşturulan regresyon ağaç diyagramı Şekil 4.6'da gösterilmiştir. Karar ağacı yapısı incelendiğinde, ön sağrı genişliği (RUMPWIDTHFRONT), incik çevresi (SHANKCIRCUMFERENCE) sağrı yüksekliği (RUMPHEIGHT), cidago yüksekliği (WITHERHEIGHT), kuyruk uzunluğu (TAILLENGTH), ve cinsiyet (SEX) bağımsız değişkenlerinin canlı ağırlık üzerinde etkili olduğu belirlenmiştir (Düz. $P < 0.05$). Regresyon ağaç diyagramının en üstünde bulunan kök düğüm ön sağrı genişliği bakımından birinci ağaç derinliğinde üç alt gruba (Düğüm 1, 2 ve 3) ayrılmıştır (Düz. $P = 0,000$, $F = 28,750$, $sd1 = 2$ ve $sd2 = 161$).

Düğüm 1, ön sağrı genişliği ≤ 14 cm olan hayvanların oluşturduğu alt grubu ifade etmektedir. Düğüm 1'de bulunan hayvanlar toplam hayvan sayısının %84.8'ini oluşturmuşlardır. Düğüm 2, ön sağrı genişliği $14 \text{ cm} < \text{ön sağrı genişliği} \leq 14,50$ cm olan hayvanların oluşturduğu alt grubu ifade etmektedir. Düğüm 3, ön sağrı genişliği $> 14,50$ cm olan hayvanların oluşturduğu alt grubu ifade etmektedir.

Düğüm 4, ön sağrı yüksekliği ≤ 14 cm ve incik çevresi ≤ 8 cm olan hayvanların oluşturduğu alt grubu temsil etmektedir (21.127 kg). Düğüm 4, kuyruk uzunluğu özelliği bakımından 3 alt gruba (Düğüm 10, 11 ve 12) ayrılmaktadır (Düz. $P = 0,016$, $F = 8,334$, $sd1 = 2$ ve $sd2 = 48$). Düğüm 10, ön sağrı yüksekliği ≤ 14 cm, incik çevresi ≤ 8 cm ve kuyruk uzunluğu $\leq 12,5$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 11, ön sağrı yüksekliği ≤ 14 cm, incik çevresi ≤ 8 cm ve kuyruk uzunluğu $12,50 < \text{kuyruk uzunluğu} \leq 15,5$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 12, ön sağrı yüksekliği ≤ 14 cm, incik çevresi ≤ 8 cm ve kuyruk uzunluğu $> 15,5$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir.

Düğüm 5, ön sağrı genişliği ≤ 14 cm ve incik çevresi > 8 cm olan hayvanların oluşturduğu alt grubu temsil etmektedir (Düz. $P = 0,003$, $F = 14,281$, $sd1 = 1$ ve $sd2 = 86$). Düğüm 5, sağrı yüksekliği bakımından iki alt gruba ayrılmıştır (Düğüm 13 ve 14). Düğüm 13, ön sağrı genişliği ≤ 14 cm, incik çevresi > 8 cm ve sağrı yüksekliği $\leq 59,50$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 14, ön sağrı genişliği ≤ 14 cm, incik

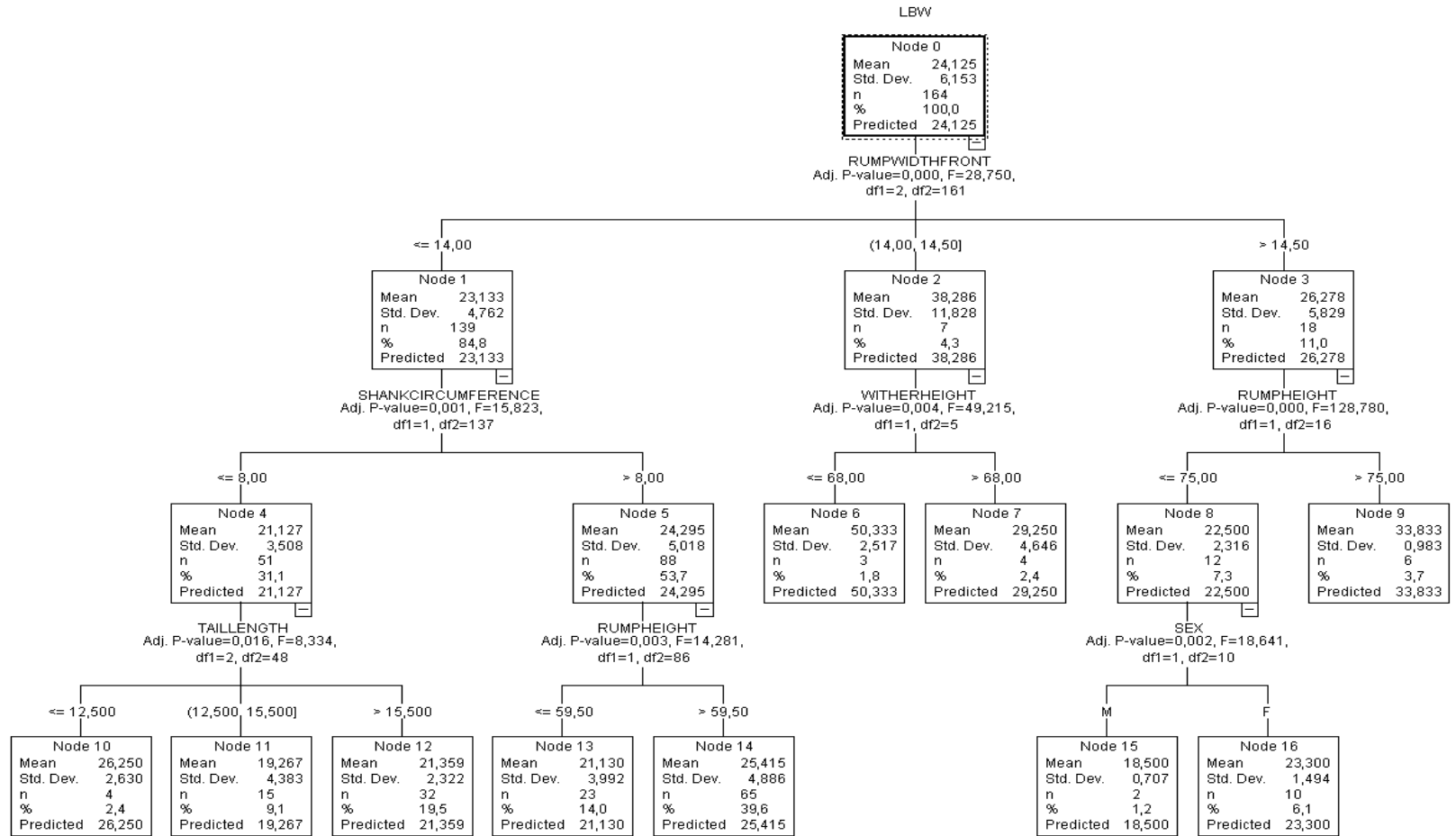
çevresi >8 cm ve sağrı yüksekliği >59,50 cm olan hayvanların oluşturduğu alt grubu temsil etmektedir.

Düğüm 6, ön sağrı genişliği $14 \leq$ ön sağrı genişliği $\leq 14,50$ cm ve cidago yüksekliği < 68 cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 6'yı oluşturan hayvanlar yeterince homejen bir yapıya ulaştığı için herhangi bir bölünmeye uğramamıştır.

Düğüm 7, $14 <$ ön sağrı genişliği $\leq 14,50$ cm ve cidago yüksekliği > 68 cm olan hayvanların oluşturduğu alt grubu temsil etmektedir.

Düğüm 8, ön sağrı genişliği > 14,5 cm, sağrı yüksekliği ≤ 75 cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 8, cinsiyet özelliği bakımından iki alt gruba ayrılmıştır (Düğüm 15 ve 16). Düğüm 15, ön sağrı genişliği > 14.5 cm, sağrı yüksekliği ≤ 75 cm ve cinsiyeti erkek olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 16, ön sağrı genişliği > 14.5 cm, sağrı yüksekliği ≤ 75 cm ve cinsiyeti dişi olan hayvanların oluşturduğu alt grubu temsil etmektedir.

Düğüm 9, ön sağrı genişliği > 14 cm ve sağrı yüksekliği > 75 cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 9'u oluşturan hayvanlar yeterince homojen bir yapıya ulaştığı için 3. Ağaç denkleğinde bölünme işlemi gerçekleşmemiştir.

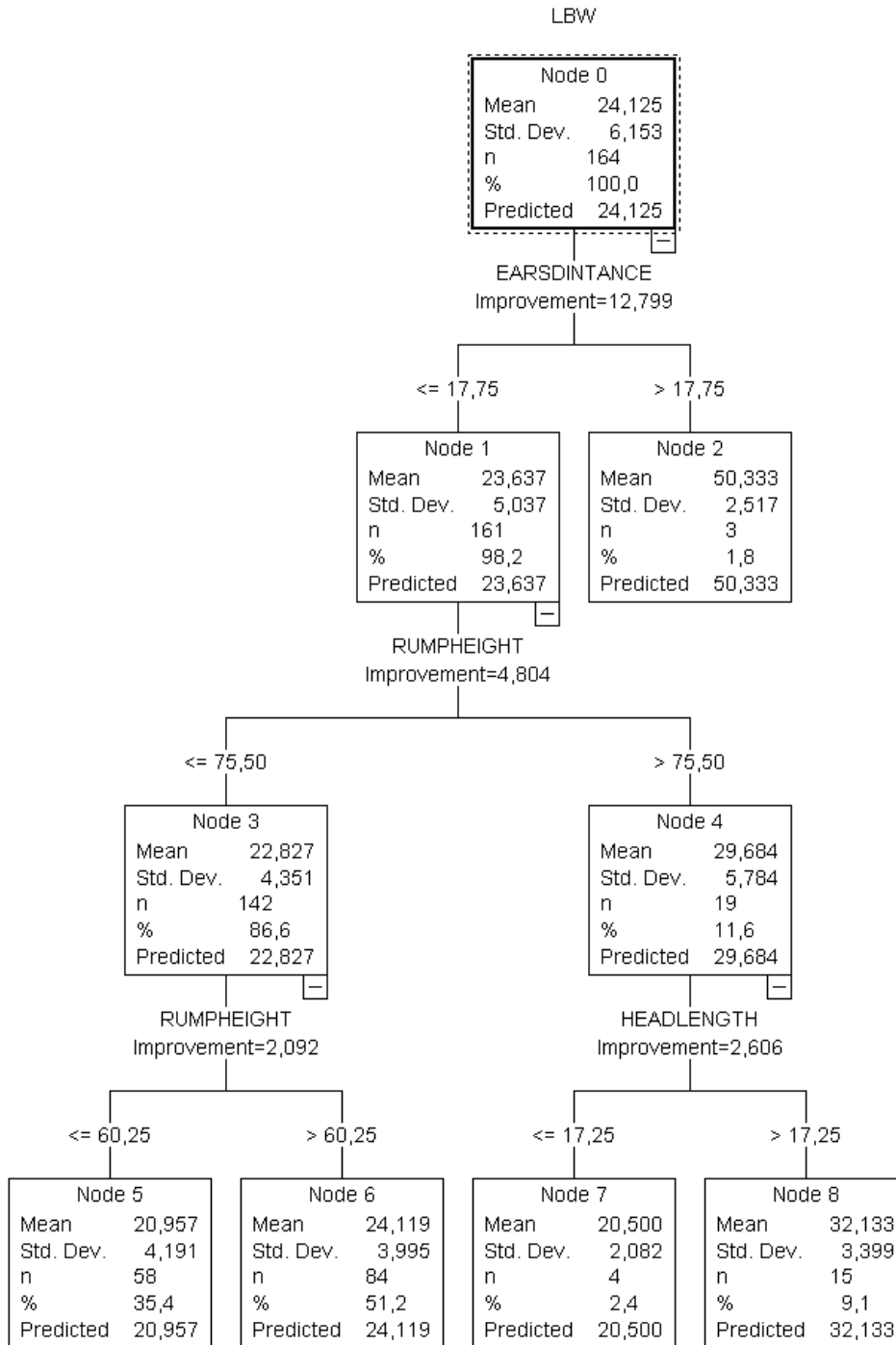


Şekil 4.6. CHAID algoritması ile oluşturulan ağaç diyagramı

4.4. CART algoritmasına ilişkin sonuçlar

CART algoritmasıyla oluşturulan regresyon ağaç diyagramı Şekil 4.7’de gösterilmiştir. Ağaç yapısı incelendiğinde, kulaklararası mesafe (EARSDISTANCE), sağrı yüksekliği (RUMPHEIGHT) ve baş uzunluğu (HEADLENGTH) bağımsız değişkenlerinin canlı ağırlık üzerinde etkili olduğu belirlenmiştir. Regresyon ağaç diyagramının en üstünde bulunan kök düğüm 1. Ağaç derinliğinde kulaklararası mesafe bakımından iki alt gruba (Düğüm 1 ve 2) ayrılmıştır.

Düğüm 1, kulaklararası mesafe $\leq 17,75$ cm olan hayvanların oluşturduğu alt grubu ifade etmektedir. Düğüm 1’de bulunan hayvanlar toplam hayvan sayısının % 98.2’ini oluşturmuşlardır. Düğüm 1, sağrı yüksekliği bakımından iki alt gruba ayrılmıştır (Düğüm 3 ve 4). Düğüm 3, kulaklararası mesafe $\leq 17,75$ cm ve sağrı yüksekliği $\leq 75,50$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 3, sağrı yüksekliği bakımından iki alt gruba (Düğüm 5 ve 6) ayrılmıştır. Düğüm 5, kulaklararası mesafe $\leq 17,75$ cm ve sağrı yüksekliği $\leq 60,25$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 6, kulaklararası mesafe $\leq 17,75$ cm ve sağrı yüksekliği $> 60,25$ cm olan hayvanların oluşturduğu alt grubu ifade etmektedir. Düğüm 4, kulaklararası mesafe $\leq 17,75$ cm ve sağrı yüksekliği $> 75,50$ cm olan hayvanların oluşturduğu alt grubu ifade etmektedir. Düğüm 4, baş uzunluğu bakımından iki alt gruba (Düğüm 7 ve 8) ayrılmıştır. Düğüm 7, kulaklararası mesafe $\leq 17,75$ cm, sağrı yüksekliği $> 75,50$ cm ve baş uzunluğu $\leq 17,25$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir. Düğüm 8, kulaklararası mesafe $\leq 17,75$ cm, sağrı yüksekliği $> 75,50$ cm ve baş uzunluğu $> 17,25$ cm olan hayvanların oluşturduğu alt grubu temsil etmektedir.



Şekil 4.7. CART algoritması ile oluşturulan ağaç diyagramı

4.5. MARS Algoritmasına İlişkin Sonuçları

MARS algoritması ile oluşturulan tahmin denkleminde ilişkin temel fonksiyonlar ve katsayılar Çizelge 4.2’de verilmiştir. En düşük GCV (2.2) değerine karşılık gelen 45 terimlik MARS tahmin modelinde, belirleme sayısı 0.941 R^2 olarak hesaplanmıştır. Söz konusu bu tahmin denkleminde, canlı ağırlık özelliğine ilişkin toplam farklılığın % 94.1’i modele dahil edilen terimler tarafından açıklanmıştır. Modelde bulunan katsayıların tamamı istatistiksel olarak önemli bulunmuştur ($P < 0.01$). Modele dahil edilen bağımsız değişkenlere ait relatif önem dereceleri sırasıyla, sağrı yüksekliği (RUMPHEIGHT, 21/45), boynuz uzunluğu (HORNLENGTH, 15/45), boyun çevresi (NECKGIRTH, 13/45), kafa uzunluğu (HEADLENGTH, 8/45), arka sağrı genişliği (RUMPWIDTHBACK, 8/45), göğüs çevresi (HEARTGIRTH, 8/45), kuyruk uzunluğu (TAILLENGTH, 5/45), sağrı uzunluğu (RUMPLENGTH, 5/45), cidago yüksekliği (WITHERSHEIGHT, 4/45), ağız genişliği (MOUTHWIDTH, 3/45), vücut derinliği (BODYDEPTHHEART, 3/45), incik çevresi (SHANKCIRUM, 3/45), karın çevresi (BELLYGIRTH, 3/45), ön sağrı genişliği (RUMPWIDTHFRONT, 3/45 ve vücut uzunluğu (BODYLENGTH, 3/45) olarak bulunmuştur.

Çizelge 4.2. MARS tahmin denkleminde ait terimler ve katsayıları

Temel Fonksiyonlar	Katsayılar
Sabit	26.361
max(0, 17.5-HEADLENGTH)	-8.696
max(0, HORNLENGTH-16.5)	9.610
max(0, 76-RUMPHEIGHT)	-0.400
max(0, RUMPHEIGHT-76)	0.888
max(0, HORNLENGTH-16.5) * RUMPWIDTHBACK	-0.141
max(0, HORNLENGTH-16.5) * WITHERHEIGHT	-0.099
max(0, HEADLENGTH-18.5) * max(0, 16.5-HORNLENGTH)	-0.262
max(0, 18-HEADLENGTH) * max(0, 76-RUMPHEIGHT)	0.350
max(0, HEADLENGTH-18) * max(0, 76-RUMPHEIGHT)	0.060
max(0, 12-MOUTHWIDTH) * max(0, 69-HEARTGIRTH)	-11.183
max(0, 14-HORNLENGTH) * max(0, 14-RUMPLENGTH)	-0.505
max(0, NECKGIRTH-30) * max(0, 76-RUMPHEIGHT)	-0.388
max(0, 58-BODYLENGTH) * max(0, 14-RUMPLENGTH)	-0.702
max(0, HEARTGIRTH-69) * max(0, 5.5-RUMPWIDTHBACK)	0.712
max(0, HEARTGIRTH-66) * max(0, 16-TAILLENGTH)	-0.288
max(0, HEARTGIRTH-69) * max(0, TAILLENGTH-16.5)	0.115
max(0, HEARTGIRTH-69) * max(0, 16.5-TAILLENGTH)	0.387
max(0, 69-HEARTGIRTH) * max(0, WITHERHEIGHT-63.5)	0.036
max(0, 69-HEARTGIRTH) * max(0, 63.5-WITHERHEIGHT)	0.246
max(0, BODYDEPTHHEART-27) * max(0, 14-RUMPLENGTH)	1.306
max(0, 27-BODYDEPTHHEART) * max(0, 76-RUMPHEIGHT)	0.117
max(0, 13.4-TAILLENGTH) * max(0, 76-RUMPHEIGHT)	0.174
max(0, SHANKCIRCUM-9) * max(0, 76-RUMPHEIGHT)	-0.216
max(0, 12-MOUTHWIDTH) * NECKGIRTH * max(0, 69-HEARTGIRTH)	0.396
max(0, HORNLENGTH-16.5) * RUMPWIDTHBACK * max(0, 14-TAILLENGTH)	-0.057
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, NECKGIRTH-29)	-0.286
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, 29-NECKGIRTH)	-0.239
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, BELLYGIRTH-73)	-0.064
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, RUMPWIDTHFRONT-12)	0.306
max(0, MOUTHWIDTH-12.2) * max(0, 30-NECKGIRTH) * max(0, 76-RUMPHEIGHT)	-0.165
max(0, HORNLENGTH-13) * max(0, 30-NECKGIRTH) * max(0, 76-RUMPHEIGHT)	-0.023
max(0, 16.5-HORNLENGTH) * max(0, 75-BELLYGIRTH) * max(0, RUMPWIDTHBACK-4.5)	0.027
max(0, 16.5-HORNLENGTH) * max(0, RUMPWIDTHBACK-4.5) * max(0, 64-WITHERHEIGHT)	0.028
max(0, 16.5-HORNLENGTH) * max(0, 14-RUMPLENGTH) * max(0, RUMPHEIGHT-59)	-2.379
max(0, 16.5-HORNLENGTH) * max(0, 9-SHANKCIRCUM) * max(0, RUMPHEIGHT-59)	0.104
max(0, 30-NECKGIRTH) * max(0, BODYLENGTH-59) * max(0, 76-RUMPHEIGHT)	0.039
max(0, 30-NECKGIRTH) * max(0, 59-BODYLENGTH) * max(0, 76-RUMPHEIGHT)	0.014
max(0, 30-NECKGIRTH) * max(0, RUMPWIDTHFRONT-12) * max(0, 76-RUMPHEIGHT)	-0.086
max(0, 30-NECKGIRTH) * max(0, 12-RUMPWIDTHFRONT) * max(0, 76-RUMPHEIGHT)	-0.096
max(0, NECKGIRTH-30) * max(0, RUMPWIDTHBACK-4.5) * max(0, 76-RUMPHEIGHT)	0.093
max(0, NECKGIRTH-30) * max(0, RUMPWIDTHBACK-8) * max(0, 76-RUMPHEIGHT)	-0.088
max(0, 76-BELLYGIRTH) * max(0, 9-SHANKCIRCUM) * max(0, 76-RUMPHEIGHT)	-0.034
max(0, 16.5-HORNLENGTH) * BODYDEPTHHEART * max(0, 14-RUMPLENGTH) * max(0, RUMPHEIGHT-59)	0.095
max(0, NECKGIRTH-30) * max(0, 4.5-RUMPWIDTHBACK) * SHANKCIRCUM * max(0, 76-RUMPHEIGHT)	0.053

Relatif önem derecesi en yüksek olan sağrı yüksekliği özelliğinden yola çıkarak MARS tahmin denkleminin terim sayısını azaltmak için sağrı yüksekliği 76 cm ve daha yüksek olduğu düşünüldüğünde 17 adet terimin etkisi maskelenmiş olacaktır. MARS II tahmin denkleminin terimleri ve katsayıları Çizelge 4.3.'de verilmiştir.

Çizelge 4.3. MARS maskelenmiş tahmin denklemi II'e ait terimler ve katsayıları

Temel Fonksiyonlar	Katsayılar
Sabit	26.361
max(0, 17.5-HEADLENGTH)	-8.696
max(0, HORNLENGTH-16.5)	9.610
max(0, HORNLENGTH-16.5) * RUMPWIDTHBACK	-0.141
max(0, HORNLENGTH-16.5) * WITHERHEIGHT	-0.099
max(0, HEADLENGTH-18.5) * max(0, 16.5-HORNLENGTH)	-0.262
max(0, 12-MOUTHWIDTH) * max(0, 69-HEARTGIRTH)	-11.183
max(0, 14-HORNLENGTH) * max(0, 14-RUMPLENGTH)	-0.505
max(0, 58-BODYLENGTH) * max(0, 14-RUMPLENGTH)	-0.702
max(0, HEARTGIRTH-69) * max(0, 5.5-RUMPWIDTHBACK)	0.712
max(0, HEARTGIRTH-66) * max(0, 16-TAILLENGTH)	-0.288
max(0, HEARTGIRTH-69) * max(0, TAILLENGTH-16.5)	0.115
max(0, HEARTGIRTH-69) * max(0, 16.5-TAILLENGTH)	0.387
max(0, 69-HEARTGIRTH) * max(0, WITHERHEIGHT-63.5)	0.036
max(0, 69-HEARTGIRTH) * max(0, 63.5-WITHERHEIGHT)	0.246
max(0, BODYDEPTHHEART-27) * max(0, 14-RUMPLENGTH)	1.306
max(0, 12-MOUTHWIDTH) * NECKGIRTH * max(0, 69-HEARTGIRTH)	0.396
max(0, HORNLENGTH-16.5) * RUMPWIDTHBACK * max(0, 14-TAILLENGTH)	-0.057
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, NECKGIRTH-29)	-0.286
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, 29-NECKGIRTH)	-0.239
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, BELLYGIRTH-73)	-0.064
max(0, 18.5-HEADLENGTH) * max(0, 16.5-HORNLENGTH) * max(0, RUMPWIDTHFRONT-12)	0.306
max(0, 16.5-HORNLENGTH) * max(0, 75-BELLYGIRTH) * max(0, RUMPWIDTHBACK-4.5)	0.027
max(0, 16.5-HORNLENGTH) * max(0, RUMPWIDTHBACK-4.5) * max(0, 64-WITHERHEIGHT)	0.028
max(0, 16.5-HORNLENGTH) * max(0, 14-RUMPLENGTH) * max(0, RUMPHEIGHT-59)	-2.379
max(0, 16.5-HORNLENGTH) * max(0, 9-SHANKCIRCUM) * max(0, RUMPHEIGHT-59)	0.104
max(0, 16.5-HORNLENGTH) * BODYDEPTHHEART * max(0, 14-RUMPLENGTH) * max(0, RUMPHEIGHT-59)	0.095

MARS tahmin denkleminin terim sayısını azaltmak için sağrı yüksekliği ≥ 76 cm ve boynuz uzunluğu ≥ 16.5 cm olan hayvanlar düşünüldüğünde MARS tahmin modeli sadece 12 adet terimden oluşacaktır. MARS III tahmin denkleminin terimleri ve katsayıları Çizelge 4.4.'de verilmiştir.

Çizelge 4.4. MARS maskelenmiş tahmin denklemi III'e ait terimler ve katsayıları

Temel Fonksiyonlar	Katsayılar
Sabit	26.361
$\max(0, 17.5\text{-HEADLENGTH})$	-8.696
$\max(0, 12\text{-MOUTHWIDTH}) * \max(0, 69\text{-HEARTGIRTH})$	-11.183
$\max(0, 58\text{-BODYLENGTH}) * \max(0, 14\text{-RUMPLENGTH})$	-0.702
$\max(0, \text{HEARTGIRTH-69}) * \max(0, 5.5\text{-RUMPWIDTHBACK})$	0.712
$\max(0, \text{HEARTGIRTH-66}) * \max(0, 16\text{-TAILLENGTH})$	-0.288
$\max(0, \text{HEARTGIRTH-69}) * \max(0, \text{TAILLENGTH-16.5})$	0.115
$\max(0, \text{HEARTGIRTH-69}) * \max(0, 16.5\text{-TAILLENGTH})$	0.387
$\max(0, 69\text{-HEARTGIRTH}) * \max(0, \text{WITHERHEIGHT-63.5})$	0.036
$\max(0, 69\text{-HEARTGIRTH}) * \max(0, 63.5\text{-WITHERHEIGHT})$	0.246
$\max(0, \text{BODYDEPTHHEART-27}) * \max(0, 14\text{-RUMPLENGTH})$	1.306
$\max(0, 12\text{-MOUTHWIDTH}) * \text{NECKGIRTH} * \max(0, 69\text{-HEARTGIRTH})$	0.396

MARS tahmin denkleminin terim sayısını azaltmak için sağrı yüksekliği ≥ 76 cm, boynuz uzunluğu ≥ 16.5 cm ve göğüs çevresi ≥ 69 cm olan hayvanlar düşünüldüğünde MARS tahmin modeli sadece 5 adet terimden oluşacaktır. MARS IV tahmin denklemine ait terimler ve katsayılar Çizelge 4.5'de verilmiştir.

Çizelge 4.5. MARS maskelenmiş tahmin denklemi IV'e ait terimler ve katsayıları

Temel Fonksiyonlar	Katsayılar
Sabit	26.361
$\max(0, 17.5\text{-HEADLENGTH})$	-8.696
$\max(0, 58\text{-BODYLENGTH}) * \max(0, 14\text{-RUMPLENGTH})$	-0.702
$\max(0, \text{HEARTGIRTH-66}) * \max(0, 16\text{-TAILLENGTH})$	-0.288
$\max(0, \text{BODYDEPTHHEART-27}) * \max(0, 14\text{-RUMPLENGTH})$	1.306

$$CA = 26.361 - 8.696 * \max(0, 17.5\text{-HEADLENGTH}) - 0.702 * \max(0, 58\text{-BODYLENGTH}) * \max(0, 14\text{-RUMPLENGTH}) - 0.288 * \max(0, \text{HEARTGIRTH-66}) * \max(0, 16\text{-TAILLENGTH}) + 1.306 * \max(0, \text{BODYDEPTHHEART-27}) * \max(0, 14\text{-RUMPLENGTH})$$

Mevcut bu çalışmada, canlı ağırlığın tahmin edilmesi bakımından CART, CHAID ve Exhaustive CHAID algoritmalarına nazaran MARS algoritmasının üstün tahmin performans gösterdiği tespit edilmiştir. Eyduran (2017a), Beetal keçileri üzerinde yaptıkları çalışmada, OLS, CART, CHAID, RBF, MLP1 ve MLP2 algoritmalarına ilişkin standart

sapma oranlarını kabul edilebilir sınırın (0.40) üzerinde bulmuştur. Bu durum, MARS algoritması dışında CART, CHAID ve Exhaustive CHAID algoritması içinde geçerli olmuştur (Çizelge 4.1). Bunun nedeni, çalışılan sürü büyüklüğünün yetersiz olmasından kaynaklanabilir. Ancak, yukarıda değinilen diğer çalışmalarda olduğu gibi sürü büyüklüğünün yetersiz olduğu durumlarda bile MARS algoritmasının önemli değişkenlerin belirlemedeki üstünlüğü dikkat çekmektedir. Bir yaşlı Pakistan yerli koyunlarının canlı ağırlık özelliğini tahmin etmek amacıyla 26 ve 38 terimli MARS algoritması kullanan Akkus (2017), çok yüksek bir tahmin performansı elde ettiğini belirtmiştir. Ayrıca, Celik et al. (2017a) Mengali koçların canlı ağırlıklarını tahmin etmek amacıyla oluşturulan interaksyonlu ve interaksyonsuz MARS tahmin denklemlerinin canlı ağırlığa ait toplam farklılığın yaklaşık %90'ını açıkladığını göstermiştir. Farklı hayvan materyalleri ile yapılan çalışmalar da incelendiğinde, MARS algoritmasına ilişkin yapılan belirleme katsayısı tahminleri ile mevcut çalışma sonuçları birbirini destekler niteliktedir (Eyduvan *et al.*, 2017b; Aytekin *et al.*, 2018).

Doğu Anadolu'da yetiştirilen yerli, melez ve kültür ırkı sığırlarda besi performansını etkileyen faktörleri belirlemek amacıyla yürütülen önceki anket çalışmalarında interaksyon terimleri içeren MARS algoritmasının oldukça yüksek tahmin performansı ürettiği belirtilmiştir (Ertürk, 2018; Aksoy et al., 2018a; Ertürk et al., 2018). Aksoy *et al.* (2018c) yarı entansif koşullarda yetiştirilen Morkaraman ve Romanov x Morkaraman F₁ melezi kuzularından elde edilen mera gelirini tahmin etmek amacıyla kullanılan MARS veri madenciliği algoritması ile kurulan modelin tahmin doğruluğunun % 100'e yakın olduğunu bildirmişlerdir.

Yakubu (2012) Uda koçları üzerinde yürüttüğü çalışmada CART algoritması ile canlı ağırlığa ait toplam farklılığın yaklaşık %62'sinin açıklandığını belirlemiştir. Elde edilen bu sonuç, mevcut çalışma ile paralellik göstermektedir. Buna karşın, Nijerya Muscovy ördekleri üzerinde çalışma yürüten Oguntunji (2017), canlı ağırlık tahmin etmek amacıyla kullanmış olduğu CART algoritmasından yüksek tahmin performansı (0.91 R²) elde etmiştir. Benzer şekilde, Mendeş and Akkartal (2009) Ross-308 etçi tavukların canlı ağırlık özelliğini tahmin etmek amacıyla kullanmış oldukları CART algoritmasının iyi

performans ($0.81 R^2$) gösterdiğini ifade etmiştir. Ancak, Oguntunji (2017) ve Mendes and Akkartal (2009) bu çalışmada kullanılan diğer uyum iyiliği ölçütlerini rapor etmemişlerdir.

Mohammad *et al.* (2012), Pakistan'da yetiştirilen Mengali, Balochi, Harnai, Beverigh ve Rakshani ırkları üzerinde yaptıkları çalışmada CHAID algoritması kullanarak canlı ağırlıktaki toplam varyasyonun %72'sinin açıklandığını belirtmişlerdir. Ali et al. (2015) tarafından yürütülen çalışmada, Harnai kuzularının canlı ağırlık özelliği tahmin edilmesi bakımından CART, CHAID, Exhaustive CHAID ve ANN algoritmalarının tahmin performansları değerlendirilmiş, canlı ağırlık bakımından açıklanan varyasyon oranının %80'in üzerinde olduğu ve standart sapma oranlarının kabul edilebilir sınırlara yakın olduğu tespit edilmiştir. Khan *et al.* (2014), Harnai koyunlarından alınan vücut ölçülerinden yola çıkarak canlı ağırlığı tahmin etmek amacıyla Exhaustive CHAID algoritmasını kullanmış, canlı ağırlığa ait toplam açıklanan varyasyonun % 84.4 (R^2) olduğunu belirtmişlerdir. Balta and Topal (2018), Hemşin kuzuları üzerinde yürüttükleri çalışmada, çevre faktörlerinin canlı ağırlık üzerindeki etkilerini tespit etmeye çalışmış, açıklanan varyans oranını % 86,2 (R^2) olarak bulmuştur. Mengali kuzularının sütten kesim ağırlığını etkileyen çevre faktörlerini belirlemek amacıyla yapılan önceki bir çalışmada, Koc et al. (2017) CART, CHAID ve Exhaustive CHAID algoritmalarını tahmin performansı bakımından hemen hemen birbirine yakın sonuçlar ($0.70-0.75 R^2$) elde etmişler, algoritmalara ait standart sapma oranlarının kabul edilebilir sınırın üzerinde olduğunu ve CHAID algoritması ile üretilen regresyon ağaç yapısının biyolojik anlamda daha uygun olabileceğini rapor etmişlerdir. Karabacak *et al.* (2017), Konya ilinde yetiştirilen Akkaraman, Dağlıç, Kivircik, Merinos ve Karacabey Merinosu koyunlarında canlı ağırlık tahmin performansları bakımından CHAID ve CART algoritmalarının birbirine benzer sonuçlar ürettiği, CHAID algoritmasının biyolojik anlamda daha uygun ağaç yapısı ürettiğini ifade etmişlerdir.

Celik *et al.* (2017a) Mengali koçlarından alınan bazı testis ve vücut özelliklerinden canlı ağırlığı tahmin etmek amacıyla kullanmış oldukları algoritmaların tahmin performans sıralamasının $CART > CHAID \approx Exhaustive CHAID > MARS_2 > MARS_1 > RBF > MLP$ şeklinde olduğunu ifade etmişlerdir. Mevcut çalışmanın aksine Olfaz et al. (2018)

Karayaka kuzuları üzerinde yürüttükleri çalışmada, CART ve CHAID algoritmalarına ait % 88 R^2 ve kabul edilebilir standart sapma oranları (≤ 0.40) bulduklarını belirtmişlerdir.

Eyduran (2016) küçükbaş hayvanlarda canlı ağırlığın tahmin edilmesi kapsamında çoklu regresyonda bazı temel varsayımların ihlal edilmesi ve çoklu bağlantılılığın olduğu durumlarda CART, CHAID, Exhaustive CHAID, MLP ve MARS gibi veri madenciliği algoritmalarının çoklu regresyon analizine alternatif olacağını vurgulamıştır. Akkus (2017) ve Aksoy et al. (2018b) çoklu regresyon sonuçlarının, MARS sonuçlarına nazaran daha düşük tahmin doğruluğu ürettiğini belirtmişlerdir.

Canlı ağırlık özelliğini tahmin etmek amacıyla earth paketinde $penalty=-1$ ve $degree=2$ kısıtlamaları yapan Akkuş (2017), MARS algoritmasının oldukça iyi bir tahminleme performansı sergilediğini göstermiştir. R programında earth paketini geliştiren Millborrow (2018), $penalty=-1$ ile yüksek bir “nk” (maksimum terim sayısı) ve düşük bir “nprune (istenilen terim sayısı) tanımlaması ile iyi tahminlemeler yapılabileceğini vurgulamıştır. Ancak, bazı temel kaynaklarda, MARS veri madenciliği algoritmasının çoklu bağlantılılık problemine karşı hassas olduğu da belirtilmiştir (Friedman, 1971; Millborrow, 2018).

Benzer çalışmalar incelendiğinde, tahmin performansına ilişkin literatürdeki farklılıklar; yetiştirme sistemi, hayvanın sayısı, türü, ırkı, yaşı, doğum tipi, doğum yılı, ana yaşı, ananın cüssesi, cinsiyeti, bağımsız değişkenler, kullanılan algoritmalar ve algoritmalara ait yapılan (minimum düğüm sayısı, eğitim: test set oranı, ağaç yapısının derinliği) bazı kısıtlamalara ait farklılıklardan kaynaklanabilir.

Çizelge 4.6. Canlı ağırlığı tahmin etmek amacıyla eski yıllarda yapılan bazı çalışmalar

Çalışmalar	Algoritma	r	R ²	Düz R ²	RMSE	MAD	RAE	MAPE	CV (%)	AIC	AIC C	SS Oran
Yakubu (2012)	CART		0,62									
Uda Koçlar Mohamma d et al. (2012) (n=239)	CHAID		0,72									
Khan et al. (2014) Harnai Koyunu	E. CHAID		0,84									
Ali et al. (2015) Harnai Kuzuları	CART	0,909	0,826	0,822	1,566		0,058		5,91			0,42
	CHAID	0,915	0,838	0,833	1,509		0,056		5,71			0,40
	E. CHAID	0,918	0,842	0,838	1,488		0,056		5,63			0,40
	ANN	0,906	0,820	0,815	1,589		0,059		5,99			0,42
Eyduran et al (2017a) Beetal Keçisi	OLS	0,862			3,973	2,786	0,089	6,9		582		0,51
	CART	0,821			4,469	3,325	0,100	8,1		620		0,57
	CHAID	0,847			4,157	2,990	0,093	7,3		594		0,53
	RBF	0,864			3,940	2,888	0,088	7,0		1172		0,50
	MLP1	0,820			4,486	3,259	0,100	8,0		891		0,57
	MLP2	0,834			4,330	3,154	0,097	7,7		-30		0,55

Çizelge 4.6. (Devam)

Çalışmalar	Algoritma	r	R ²	Düz R ²	RMSE	MAD	RAE	MAPE	CV (%)	AIC	AIC C	SS Oran
Akkuş (2017) (n=191)	OLS		0,653	0,642								0,59
	MARS	0,990	0,980	0,975								0,01
Balta and Topal (2018)	CART		0,862									
Hemşin Kuzuları Eyduran et al. (2017b)	MARS	0,94	0,88	0,87								0,35
Karabacak et al. (2017)	CHAID	0,805	0,648	0,632	1,181		0,058		5,94			0,59
	E. CHAID	0,787	0,620	0,612	1,226		0,061		6,17			0,62
	CART	0,754	0,569	0,549	1,306		0,065		6,57			0,66
	MARS	0,939	0,881	0,877	4,527	3,373	0,087	7,1		333	335	0,35
Celik et al. (2017a)	MARS İnt.	0,941	0,885	0,880	4,450	3,355	0,085	7,2		330	331	0,34
	CHAID	0,949	0,901	0,899	4,144	2,991	0,080	6,5		310	311	0,31
	E. CHAID	0,949	0,901	0,899	4,138	2,980	0,079	6,5		312	313	0,31
	CART	0,959	0,920	0,917	3,704	2,664	0,071	5,8		291	291	0,28
Mengali Koçları	MLP	0,926	0,857	0,847	4,995	3,746	0,096	7,9		361	362	0,38

Çizelge 4.6. (Devam)

Çalışmalar	Algoritma	r	R ²	Düz R ²	RMSE	MAD	RAE	MAPE	CV (%)	AIC	AIC C	SS Oran
Koc et al. (2017)	CHAID	0,863	0,744	0,743	0,714		0,178		4,46			0,51
Mengali Kuzusu	E. CHAID	0,862	0,743	0,743	0,715		0,179		4,46			0,51
	CART	0,844	0,716	0,712	0,756		0,189		4,72			0,54
Aytekin et al. (2018) Kültür ırkı ve melez genç boğalar (n=103)	MARS İnt.	0,986	0,972	0,964								0,17
Oguntunji (2017) Nijerya Muscovy ördekleri	CART	0,954	0,910									
Mendeş and Akkartal (2009) Ross-308 Broyler	CART	0,900	0,810									
Olfaz et al. (2018) Karayaka kuzusu	CART	0,938	0,880		1,612	1,195						0,35
	CHAID	0,938	0,880		1,623	1,192						0,35

5. SONUÇ ve ÖNERİLER

Bu çalışmada, sürekli bir bağımlı değişkenin tahmin edilmesi kapsamında, CART, CHAID, Exhaustive CHAID ve MARS veri madenciliği algoritmaları tahminleme performansları bakımından karşılaştırılmış, algoritmaların uygulanması ile ilgili detaylara değinilmiştir. Buna ilaveten mevcut bu çalışmada, üzerinde durulan algoritmalar ilişkin uyum iyiliği ölçütlerinin hesaplanabilmesiyle ilgili yazılmış olan R script dosyaları literatüre kazandırılmıştır (EK1 ve EK2). Araştırılan algoritmaların uyum iyiliği ölçütleri (goodness of fit criteria) ve bağımlı değişkene ait gerçek-tahmin edilen değerlere ilişkin grafikler esas alındığında, MARS veri madenciliği algoritmasının gerçek verilere yakın tahmin değerleri ürettiği belirlenmiştir. Araştırmada elde edilen sonuçlar aşağıda özetlenmiştir.

1. Tahmin performansı bakımından performans üstünlük sıralaması, MARS > Exhaustive CHAID > CHAID > CART şeklinde oluşmuştur. Ancak, CART algoritmasında daha az sayıda bağımsız değişkenin önemli olması, AIC ve AICc uyum iyiliği ölçütlerinde bir miktar sapmaya yol açmıştır.
2. Ebeveyn çocuk düğümleri için öngörülen minimum hayvan sayıları 24:12 den 4:2'e doğru azaltıldığında, ağaç yapısına dayalı algoritmalarında tahmin performanslarının arttığı gözlenmiştir.
3. Test edilen algoritmalar arasında CART daha az sayıda bağımsız değişkeni önemli bulmuştur. Bağımlı değişkene ait toplam varyasyonun açıklanması kapsamında bu algoritma başarılı olamamıştır.
4. Ağaç yapısına dayalı algoritmalara nazaran MARS algoritması çok sayıda önemli bağımsız değişken ile yüksek dereceli interaksiyon terimleri oluşturarak tahmin performansı bakımından üstünlük göstermiştir.
5. MARS algoritmasının çoklu bağlantılığa karşı hassas olduğu bildirildiğinden dolayı, bağımsız değişkenler arasındaki 0.85'den daha yüksek korelasyon katsayıları olup olmadığı incelenmiş, yüksek korelasyona sahip değişkenler veri setinden çıkarılmıştır.

6. Bu çalışmada önemli bulunan bağımsız değişkenler arasındaki ilişkiler, ileride yapılacak seleksiyon çalışmaları için yararlı olacaktır. Tahmin performansı yüksek olan MARS modeline dahil edilen bağımsız değişkenlerin önem derecesi sıralaması; sağrı yüksekliği (RUMPHEIGHT, 21/45), boynuz uzunluğu (HORNLENGTH, 15/45), boyun çevresi (NECKGIRTH, 13/45), kafa uzunluğu (HEADLENGTH, 8/45), arka sağrı genişliği (RUMPWITHBACK, 8/45), göğüs çevresi (HEARTGIRTH, 8/45), kuyruk uzunluğu (TAILLENGTH, 5/45), sağrı uzunluğu (RUMPLENGTH, 5/45), cidago yüksekliği (WITHERSHEIGHT, 4/45), ağız genişliği (MOUTHWIDTH, 3/45), vücut derinliği (BODYDEPTHHEART, 3/45), incik çevresi (SHANKCIRUM, 3/45), karın çevresi (BELLYGIRTH, 3/45), ön sağrı genişliği (RUMPWIDTHFRONT, 3/45 ve vücut uzunluğu (BODYLENGTH, 3/45) olarak bulunmuştur. İncelenen Teddy ırkında yapılacak seleksiyon çalışmalarında, sağrı yüksekliğinin majör bir etkiye sahip olabileceği anlaşılmıştır. Ancak, elde edilen sonuçların genelleştirilebilmesi açısından daha büyük sürülerde çalışılması önerilebilir.

Benzer konularda ve ırk standartlarının belirlenmesi ile ilgili gelecekte yapılması planlanan çalışmalarda;

1. Daha büyük hayvan sürülerinde çalışılması,
2. Özelliklerin hassas ölçülmesi,
3. Daha farklı tür ve ırklarda da çalışılması,
4. Daha etkili bağımsız değişkenlerin seçilmesi,
5. Koyun ve keçilerde akrabalılık (inbreeding) durumunun yani pedigrilerinin incelenmesi,
6. Çoklu bağlantılık durumunun belirlenmesi ve yüksek korelasyon gösteren bağımsız değişkenlerden birinin çıkarılması,
7. Klasik yaklaşımlardan ziyade ANNs ve MARS gibi daha etkin istatistiksel yaklaşımların kullanılması,
8. MARS algoritmasında, en düşük GCV değerini elde edecek şekilde interaksyon derecesinin kademeli olarak artırılması,

9. MARS algoritmasında $\text{penalty}=2$ yerine az terim sayılı model üretmek için $\text{penalty}=-1$ ile büyük “nk” ve küçük “nprune” tanımlaması yapılarak tahmin doğruluğunun artırılması,
10. Bağımlı değişken bakımından ölçülen ve tahmin edilen değerlerin grafiksel olarak incelenmesi,
11. Ayrıca etkili çevre faktörlerinin de göz önünde bulundurulması önerilebilir.

Sonuç olarak bu çalışma ile literatüre kazandırılan R script dosyalarının ileride yapılacak çalışmalara katkı sağlayacağı ümit edilmektedir.

KAYNAKLAR

- Aksoy, A., Ertürk, Y.E., Eyduran, E., Tariq, M.M., (2018a). Comparing predictive performances of MARS and CHAID algorithms for defining factors affecting final fattening live weight in cultural beef cattle enterprises. *Pakistan Journal of Zoology*, 50(6), 2279-2286.
- Akşahan, R. (2015). **Kimi Kültür Irkı Sığırlarda Besi Sonu Canlı Ağırlığını Etkileyen Çeşitli Vücut Ölçülerinin Regresyon Ağacı Yöntemiyle Belirlenmesi**. Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya. 27.
- Aksoy, A., Ertürk, Y.E., Erdoğan, S., Eyduran, E., Tariq, M.M., (2018b). Estimation of Honey Production in Beekeeping Enterprises from Eastern Part of Turkey through Some Data Mining Algorithms. *Pakistan Journal of Zoology*, 50(6), 2199-2207.
- Aksoy, A., Ertürk, Y.E., Eyduran, E., Tariq, M.M., (2018c). Utility of MARS algorithm for describing non-genetic factors affecting pasture revenue of Morkaraman breed and Romanov × Morkaraman F₁ Crossbred sheep under semi intensive conditions. *Pakistan Journal of Zoology*, 51(1), 235-240
- Ali, M., Eyduran, E., Tariq, M. M., Tirink, C., Abbas, F., Bajwa, M. A., Baloch, M. H., Nizamani, A. H., Waheed, A., Awan, M. A., Shah, S. H., Ahmad, Z., Jan, S., (2015). Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai Sheep. *Pakistan Journal of Zoology*, 47, 1579-1585.
- Aytekin, İ., Eyduran, E., Karadas, K., Akşahan, R., Keskin, İ., (2018). Prediction of Fattening Final Live Weight from some Body Measurements and Fattening Period in Young Bulls of Crossbred and Exotic Breeds using MARS Data Mining Algorithm. *Pakistan Journal of Zoology*, 50(1), 189-195.
- Bakır, G., Keskin, S., Mirtağoğlu, H., (2009). Evaluating the relationship between mature age milk yield and several traits using CHAID analysis in Brown Swiss Cows. *Journal of Animal and Veterinary Advances*, 8 (3), 587-589.

- Bakır, G., Keskin, S., Mirtağoğlu, H., (2010). Determination of the effective factors for 305 days milk yield by regression tree (RT) method. *Journal of Animal and Veterinary Advances*, 9 (1), 55-59.
- Balta, B., Topal, M., (2018). Regression tree approach for assessing the effects of non-genetic factors on birth weight of Hemşin lamb. *Alnteri Journal of Agricultural Science*, 33(1), 65-73.
- Celik, S., Yilmaz, O., (2017). Comparison of different data mining algorithms for prediction of Body weight from several morphological measurements in dogs. *Journal of Animal and Plant Sciences*, 27(1), 57-64.
- Celik, S. Yilmaz, O., (2018). Prediction of Body Weight of Turkish Tazi Dogs using Data Mining Techniques: Classification and Regression Tree (CART) and Multivariate Adaptive Regression Splines (MARS). *Pakistan Journal of Zoology*, 50(2), 575-583.
- Celik, S., Eydurani, E., Karadas, K., Tariq, M.M., (2017a). Comparison of predictive performance of data mining algorithms in predicting body weight in Mengali rams of Pakistan. *Revista Brasileira de Zootecnia*, 46(11), 863-872.
- Celik, S., Sengul, T., Inci, H., Sogut, B., Sengul, A.Y., Kuzu, C., Ayasan, T. (2017b). Estimation of egg weight from some external and internal quality characteristics in quail by using various data mining algorithms. *Indian Journal of Animal Sciences*, 87(12), 1524–1530.
- Celik, S., Eydurani, E., Tatliyer, A., Karadas, K., Kara, M.K., Waheed, A., (2018). Comparing predictive performances of some nonlinear functions and multivariate adaptive regression splines (mars) for describing the growth of Daera Din Panah (DDP) goat in Pakistan. *Pakistan Journal of Zoology*, 50(3), 1187-1190.
- Doğan, I., (2003). Holştayn ırkı ineklerde Süt verimine etki eden faktörlerin CHAID analizi ile incelenmesi. *Ankara Üniversitesi Veteriner Fakültesi Dergisi*, 50, 65-70.

- Duru, M., Duru, A., Karadaş, K., Eyduran, E., Cinli, H., Tariq, M.M., (2017). Effect of Carrot (*Daucus carota*) Leaf Powder on External and Internal Egg Characteristics of Hy-Line White Laying Hens. *Pakistan Journal of Zoology*, 49(1), 129-137.
- Eyduran, E., Karakus, K., Keskin, S., Cengiz, F., (2008). Determination of factors influencing birth weight using regression tree (RT) method. *Journal of Applied Animal Research*. 34, 109-112.
- Eyduran, E., Yılmaz, I., Kaygısız, A., Aktaş, Z. M., (2013b). An investigation on relationship between lactation milk yield, somatic cell count and udder traits in first lactation Turkish Saanen goat using different statistical techniques. *The Journal Animal Plant Science*, 23(4), 956-963.
- Eyduran, E., Yılmaz, I., Tariq, M. M., Kaygısız, A., (2013c). Estimation of 305-d milk yield using regression tree method in Brown Swiss cattle. *The Journal Animal Plant Science*, 23(3), 731-735.
- Eyduran, E. (2016). The possibility of using data mining algorithms in prediction of live body weights of small ruminants. *Journal of Biomedical Science*, 1, 1-4.
- Eyduran, E., Keskin, I., Ertürk, Y.E., Dağ, B., Tatliyer, A., Tirink, C., Aksahan, R., Tariq, M.M., (2016). Prediction of Fleece Weight from Wool Characteristics of Sheep Using Regression Tree Method (CHAID Algorithm). *Pakistan Journal of Zoology*, 48(4), 957-960.
- Eyduran, E., Türkoğlu, M., (2017). Procedure of Mars Algorithm for Describing the Relationship Between Body Weight and Morphological Traits of Some Migratory Birds in Iğdır Province of Turkey. *II. International Iğdır Symposium*, Iğdır Turkey, 133.
- Eyduran, E., Zaborski, D., Waheed, A., Celik, S., Karadas, K., Grzesiak, W., (2017a). Comparison of the Predictive Capabilities of Several Data Mining Algorithms and Multiple Linear Regression in the Prediction of Body Weight by Means of Body

- Measurements in the Indigenous Beetal Goat of Pakistan. *Pakistan Journal of Zoology*, 49(1), 273-282.
- Eyduran, E., Akkus, O., Kara, M.K., Tirink, C., Tariq, M. M. (2017b). Use of Multivariate Adaptive Regression Splines (Mars) in Predicting Body Weight from Body Measurements in Mengali Rams. *International Conference on Agriculture, Food, Veterinary and Pharmacy Sciences*, Cappadocia Turkey, 415.
- Eyduran, E., Tirink, C., Karahan, A.E., Türkoğlu, M. (2017c). Prediction of an upper bound of generalized cross validation in multivariate adaptive regression splines in agricultural studies. *International Conference on Computational and Statistical Methods in Applied Sciences*, Samsun Turkey, 176.
- Eyduran, E., Tirink, C., Karahan, A.E., Türkoğlu, M., Tariq, M.M. (2017d). Comparison of predictive performances of MARS and CART algorithms through R software. *International Conference on Computational and Statistical Methods in Applied Sciences*, Samsun Turkey, 181.
- Ertürk, Y.E., (2018). Description of Factors Influencing Final Fattening Weight in Domestic Beef Cattle Breeds through MARS Algorithm. *Pakistan Journal of Zoology*, 50(5), 1731-1737.
- Ertürk, Y.E., Aksoy, A., Tariq, M.M., (2018). Effect of Selected Variables Identified by MARS on Fattening Final Live Weight of Crossbred Beef Cattle in Eastern Turkey. *Pakistan Journal of Zoology*, 50(4), 1403-1412.
- Gevrekci, Y., Takma, Ç., (2018). A comparative study for egg production in layers by decision tree analysis. *Pakistan Journal of Zoology*, 50(2), 437-444.
- Grzesiak, W., Lacroix, R., Wójcik, J., Blaszczyk, P., (2003). A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science*, 83(2), 307-310.

- Grzesiak, W., Błaszczyk, P., Lacroix, R., (2006). Methods of predicting milk yield in dairy cows predictive capabilities of wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronic in Agriculture*, 54, 69-83.
- IBM Corp. Released 2015. IBM SPSS Statistics For Windows Version 23.0, Armonk, NY: IBM Corp.
- Karabacak, A., Celik, S., Tatliyer, A., Keskin, I., Ertürk, Y.E., Eyduran, E., Javed, Y., Tariq, M.M., (2017) Estimation of Cold Carcass Weight and Body Weight from Several Body Measurements in Sheep through Various Data Mining Algorithms. *Pakistan Journal of Zoology*, 49(5), 1731-1738.
- Karadas, K., Tariq, M., Tariq, M.M., Eyduran, E., (2017a). Measuring predictive performance of data mining and artificial neural network algorithms for predicting lactation milk yield in indigenous Akkaraman sheep. *Pakistan Journal of Zoology*, 49(1), 1-7.
- Khan, MA., Tariq, M. M., Eyduran, E., Tatliyer, A., Rafeeq, M., Abbas, F., Rashid, N., Awan, M.A., Javed, K., (2014). Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. *The Journal of Animal Plant Science*. 24(1), 120-126.
- Koc, Y., Eyduran, E., Akbulut, O., (2017). Application of Regression Tree Method for Different Data from Animal Science. *Pakistan Journal of Zoology*, 49(2), 599-607.
- Mendeş, M. ve Akkartal, E., (2009). Regression tree analysis for predicting slaughter weight in broilers. *Italian Journal of Animal Science*, 8, 615-624.
- Milborrow, S., (2018). Milborrow. Derived from mda:mars by T. Hastie and R. Tibshirani. Internet url: <https://CRAN.R-project.org/package=earth> (10.10.2018).
- Mohammad, M.T., Rafeeq, M., Bajwa, M.A., Awan, M.A., Abbas, F., Waheed, A., Bukhari, F.A., Akhtar, P., (2012). Prediction of body weight from body

- measurements using regression tree (RT) method for indigenous sheep breeds in Balochistan, *Pakistan. The Journal of Animal Plant Science*, 22(1), 20-24.
- Oguntunji, A. O., (2017). Regression tree analysis for predicting body weight of Nigerian muscovy duck (*Cairina moschata*). *Genetika*, 49(2), 743-753
- Olfaz, M., Tirink, C., Önder, H. (2018). Use of CART and CHAID algorithms in Karayaka sheep breeding. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 25(1), 105-110.
- Orhan, H., Eydurhan, E., Tatliyer, A., Saygici, H., (2016). Prediction of egg weight from egg quality characteristics via ridge regression and regression tree methods. *Revista Brasileira de Zootecnia*, 45(7), 380-385.
- Oruçoğlu, O., (2011). *Holstein Irkı İneklerin 305 günlük süt verimini etkileyen çevre faktörlerinin regresyon ağacı ile belirlenmesi*. Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü. Isparta. 52.
- Şengül, T., Çelik, S., Şengül, A.Y., (2018). Bildircinlarda Göğüs Etinin Rengi ve Ph'sı Üzerine Yaş, Cinsiyet ve Canlı Ağırlığın Etkisi. *Türk Tarım ve Doğa Bilimleri Dergisi*, 5(4), 523-529.
- Tahtalı, Y., Berberoğlu, E., (2017). Regression Tree Analysis for Determination of the Effective Factors on Birth Weight in Holstein Calves. *Journal of Agricultural Faculty of Gaziosmanpaşa University*, 34(3), 166-171.
- Takma, Ç., Gevrekçi, Y., Karahan, A., Atıl, H., Çevik, M., (2017). Yumurta verimi üzerine bazı özelliklerin etkisinin regresyon ağacı analizi ile belirlenmesi. *Ege Üniversitesi Ziraat Fakültesi Dergisi*, 54 (4), 459-463.
- Yakubu, A., (2012). Application of regression tree methodology in predicting the body weight of Uda sheep. Scientific papers. *Animal Science and Biotechnologies*, 45 (2), 484-490.

Yilmaz, I., Eyduran, E., Kaygisiz, A., (2013). Determination of non-genetic factors influencing birth weight using regression tree method in Brown-Swiss cattle. *Can. Journal of Applied Pharmaceutical Science*, 1(3), 382-387.

Zaborski, D., Grzesiak, W., Szewczuk, M., Eyduran, E., Tariq, M.M., Ali, M., (2018). The use of MARS method for predicting daily body weight gains in Harnai sheep. *IX International Scientific Agriculture Symposium "Agrosym 2018"*, Jahorina Bosnia Herzegovina,



EKLER

Ağaç yapısına dayalı CART, CHAID ve Exhaustive CHAID algoritmasına ilişkin tahmin edilen değerler ile gerçek değerler "harunothers.txt" isimli not defteri dosyasına kaydedilmiştir. Uyum iyiliği ölçütlerini hesaplatmak amacıyla R script dosyası oluşturulmuştur (EK1).

EK1

```
harun=read.table("C:/harunothers.txt", header = T)
str(harun)
head(harun)
m1=harun$PredictCHAID
m2= harun$PredictExCHAID
m3= harun$PredictCART
n<- length(harun$LBW)
k1 <- 6 ## CHAID
k2 <- 6 ## Exhaustive CHAID
k3 <- 3 ## CART
pc_CHAID=cor.test(harun$LBW, m1)
pc_ExCHAID=cor.test(harun$LBW, m2)
pc_CART=cor.test(harun$LBW, m3)
pc_CHAID
pc_ExCHAID
pc_CART
pearcorr_CHAID=round(cor(harun$LBW, m1), digits = 3)
pearcorr_ExCHAID=round(cor(harun$LBW, m2), digits = 3)
pearcorr_CART=round(cor(harun$LBW, m3), digits = 3)
pearcorr_CHAID
pearcorr_ExCHAID
pearcorr_CART
error_CHAID=harun$LBW-m1
error_ExCHAID=harun$LBW-m2
error_CART=harun$LBW-m3
Coefofvariation_CHAID=round(sd(error_CHAID)*100/mean(harun$LBW), digits=2)
Coefofvariation_ExCHAID=round(sd(error_ExCHAID)*100/mean(harun$LBW),
digits=2)
Coefofvariation_CART=round(sd(error_CART)*100/mean(harun$LBW), digits=2)
Coefofvariation_CHAID
Coefofvariation_ExCHAID
Coefofvariation_CART
Sdratio_CHAID=round(sd(error_CHAID)/sd(harun$LBW), digits=3)
Sdratio_ExCHAID=round(sd(error_ExCHAID)/sd(harun$LBW), digits=3)
Sdratio_CART=round(sd(error_CART)/sd(harun$LBW), digits=3)
```

Sdratio_CHAID
 Sdratio_ExCHAID
 Sdratio_CART
 $RMSE_CHAID = \text{round}(\sqrt{\text{mean}(\text{error_CHAID}^2)}, \text{digits}=3)$
 $RMSE_ExCHAID = \text{round}(\sqrt{\text{mean}(\text{error_ExCHAID}^2)}, \text{digits}=3)$
 $RMSE_CART = \text{round}(\sqrt{\text{mean}(\text{error_CART}^2)}, \text{digits}=3)$
 RMSE_CHAID
 RMSE_ExCHAID
 RMSE_CART
 $ME_CHAID = \text{round}(\text{mean}(\text{error_CHAID}), \text{digits}=3)$
 $ME_ExCHAID = \text{round}(\text{mean}(\text{error_ExCHAID}), \text{digits}=3)$
 $ME_CART = \text{round}(\text{mean}(\text{error_CART}), \text{digits}=3)$
 ME_CHAID
 ME_ExCHAID
 ME_CART
 $RAE_CHAID = \text{round}(\sqrt{\text{sum}(\text{error_CHAID}^2) / \text{sum}(\text{harun\$LBW}^2)}, \text{digits}=3)$
 $RAE_ExCHAID = \text{round}(\sqrt{\text{sum}(\text{error_ExCHAID}^2) / \text{sum}(\text{harun\$LBW}^2)}, \text{digits}=3)$
 $RAE_CART = \text{round}(\sqrt{\text{sum}(\text{error_CART}^2) / \text{sum}(\text{harun\$LBW}^2)}, \text{digits}=3)$
 RAE_CHAID
 RAE_ExCHAID
 RAE_CART
 $MAD_CHAID = \text{round}(\text{mean}(\text{abs}(\text{error_CHAID})), \text{digits}=3)$
 $MAD_ExCHAID = \text{round}(\text{mean}(\text{abs}(\text{error_ExCHAID})), \text{digits}=3)$
 $MAD_CART = \text{round}(\text{mean}(\text{abs}(\text{error_CART})), \text{digits}=3)$
 MAD_CHAID
 MAD_ExCHAID
 MAD_CART
 $MAPE_CHAID = \text{round}(\text{mean}(\text{abs}(\text{error_CHAID} / \text{harun\$LBW})) * 100, \text{digits}=3)$
 MAPE_CHAID ## Mean absolute percent error
 $MAPE_ExCHAID = \text{round}(\text{mean}(\text{abs}(\text{error_ExCHAID} / \text{harun\$LBW})) * 100, \text{digits}=3)$
 MAPE_ExCHAID ## Mean absolute percent error
 $MAPE_CART = \text{round}(\text{mean}(\text{abs}(\text{error_CART} / \text{harun\$LBW})) * 100, \text{digits} = 3)$
 MAPE_CART## Mean absolute percent error
 $Rsq_CHAID = \text{round}(1 - (\text{sum}(\text{error_CHAID}^2) / (\text{var}(\text{harun\$LBW}) * (n-1))), \text{digits}=3)$
 $Rsq_ExCHAID = \text{round}(1 - (\text{sum}(\text{error_ExCHAID}^2) / (\text{var}(\text{harun\$LBW}) * (n-1))), \text{digits}=3)$
 $Rsq_CART = \text{round}(1 - (\text{sum}(\text{error_CART}^2) / (\text{var}(\text{harun\$LBW}) * (n-1))), \text{digits}=3)$
 Rsq_CHAID
 Rsq_ExCHAID
 Rsq_CART
 $AdjRsq_CHAID = \text{round}(1 - ((1 - Rsq_CHAID) * (n-1) / (n-k1-1)), \text{digits}=3)$
 $AdjRsq_ExCHAID = \text{round}(1 - ((1 - Rsq_ExCHAID) * (n-1) / (n-k2-1)), \text{digits}=3)$
 $AdjRsq_CART = \text{round}(1 - ((1 - Rsq_CART) * (n-1) / (n-k3-1)), \text{digits}=3)$
 AdjRsq_CHAID
 AdjRsq_ExCHAID

AdjRsq_CART
AIC_CHAID=round(n*log(mean(error_CHAID^2), base=exp(1))+2*k1, digits=3)
AIC_ExCHAID=round(n*log(mean(error_ExCHAID^2), base=exp(1))+2*k2, digits=3)
AIC_CART=round(n*log(mean(error_CART^2), base=exp(1))+2*k3, digits=3)
AIC_CHAID
AIC_ExCHAID
AIC_CART
AICc_CHAID=round(n*log(mean(error_CHAID^2),
base=exp(1))+(2*k1)+(2*k1*(k1+1)/(n-k1-1)), digits=3)
AICc_ExCHAID=round(n*log(mean(error_ExCHAID^2),
base=exp(1))+(2*k2)+(2*k2*(k2+1)/(n-k2-1)), digits=3)
AICc_CART=round(n*log(mean(error_CART^2), base=exp(1))+(2*k3)+(2*k3*(k3+1)/(n-
k3-1)), digits=3)
AICc_CHAID
AICc_ExCHAID
AICc_CART

EK2

Mars algoritmasına ilişkin yapılan tüm hesaplamalar, aşağıdaki R script dosyası ile gerçekleştirilmiştir (EK2).

```
#####  
## MULTIVARIATE ADAPTIVE REGRESSION SPLINES ALGORITHM ##  
## R codes for Msc thesis of Harun SEVGENLER ##  
#####  
mydata=read.table("C:/harunmars.txt", header=T)  
str(mydata)  
summary(mydata1)  
mydata1 <-mydata[,-c(3,13)]  
str(mydata1)  
head(mydata1)  
install.packages('earth')  
install.packages('pastecs')  
library(pastecs)  
stat.desc(mydata1)  
library(earth)  
marsresult=earth(LBW~., data=mydata1, penalty=-1,  
                 degree=4, pmethod="backward", nk=1000,  
                 nfold=10, nprune=45, keepxy=T)  
summary(marsresult, digits=3) ## MARS output  
#####  
## Evaluation of importance of influential predictors  
evimp(marsresult)  
#####  
## Significance test of the coefficients in MARS modeling##  
bx<-model.matrix(marsresult)  
a.lm<-lm(mydata1$LBW~bx[,-1])  
summary(a.lm)  
#####  
n<-length(mydata1$LBW) ## description of sample size in the data set  
n ## sample size  
k= length(marsresult$selected.terms)  
k ## number of selected terms in the MARS predictive model  
cor.test(mydata1$LBW, predict(marsresult))  
Pearsoncorr=round(cor(mydata1$LBW, predict(marsresult)), digits = 3)  
Pearsoncorr ## Correlation coefficient  
error=mydata1$LBW-predict(marsresult) ## Residual terms  
sdratio=round(sd(error)/sd(mydata1$LBW), digits=3)  
sdratio ## Standard deviation ratio  
Coefofvariation=round(sd(error)*100/mean(mydata1$LBW), digits=2)
```

```

Coefofvariation ## Coefficient of variation
RMSE=round(sqrt(mean(error^2)), digits=3)
RMSE ## Root of mean square error
MSE=round((mean(error^2)), digits=3)
MSE ## Mean square error
ME=round(mean(error), digits=3)
ME ## Mean error
RAE=round(sqrt(sum(error^2)/sum(mydata1$LBW^2)), digits=3)
RAE ## Relative approximation error
MAPE=round(mean(abs(error/mydata1$LBW))*100, digits=4)
MAPE ## Mean absolute percent error
MAD=round(mean(abs(error)), digits = 3)
MAD ## Mean absolute deviation
Rsq=round(1-(sum(error^2)/(var(mydata1$LBW)*(n-1))), digits = 3)
Rsq ## Coefficient of determination
AdjRsq=round(1-(((1- Rsq)*(n-1)/(n-k-1))), digits=3)
AdjRsq ## Adjusted coefficient of determination
AIC=round(n*log(mean(error^2), base=exp(1))+2*k, digits=3)
AIC ## Akaike's information criterion
AICc=round(n*log(mean(error^2), base=exp(1))+2*k+(2*k*(k+1)/(n-k-1)), digits=3)
AICc ## The corrected Akaike's information criterion
plot(mydata1$LBW, predict(marsresult))
plot(marsresult) ## plot of observed responses vs. predicted responses in the response
variable
plotmo(marsresult)
shapiro.test(error) ## Homogeneity test of residuals
#####

```

ÖZGEÇMİŞ

21.12.1986 tarihinde Adana'nın Ceyhan ilçesinde doğdu. İlk, orta ve lise öğrenimini Ceyhan'da tamamladı. 2011 yılında Gazi Üniversitesi Çorum İktisadi ve İdari Bilimler Fakültesi İktisat Bölümünden mezun oldu. 2017 yılında Iğdır Üniversitesi Ziraat Fakültesi Zootekni Anabilim Dalında Yüksek Lisans eğitimine başladı.

