

Kuzularda Yetiştirme Tipinin Regresyon Ağacı Algoritmaları Tahmin Performansına Etkisi

Adile TATLİYER

Kahramanmaraş Sütçü İmam Üniversitesi Ziraat Fakültesi Zootekni Bölümü, Kahramanmaraş
<https://orcid.org/0000-0002-4239-7072>

E-posta: atatliyer@ksu.edu.tr

ÖZET

Bu araştırma, yetişirme tipi elit ve taban olan İvesi ırkı kuzuların sütten kesim ağırlıkları üzerine CART, CHAID ve Exhausted CHAID algoritmalarının tahminleme performanslarını karşılaştırmak amacıyla yapılmıştır. Bu çalışmada Osmaniye Toprakkale ilçesinde bulunan farklı yetişirme tipi uygulayan (elit ve taban) iki işletmeden elde edilen 2014-2015 yıllarında Kasım-Ocak aylarında doğmuş toplam 331 baş İvesi ırkı kuzu kaydı kullanılmıştır. Elit sürü tipinde doğan kuzuların doğum ağırlığı (DA) ortalaması 4.92 ± 0.05 kg; taban sürüde doğan kuzuların DA ortalaması ise 4.11 ± 0.07 kg; elit sürüde yetişen kuzuların sütten kesim ağırlığı (SKA) ortalaması 14.35 ± 0.12 kg; taban sürüde yetişenlerin ise 13.89 ± 0.16 kg olduğu görülmüştür. Her iki sürüde (elit ve taban) doğumdaki ana yaşı, doğum tipi, cinsiyet, doğum ayı, sütten kesim zamanı ve doğum ağırlığı bağımsız değişken olarak kullanılmıştır. CART, CHAID ve Exhausted CHAID algoritmalarını mukayese etmek için RMSE, MAPE, RAE, SD_{ratio} ve MAD uyum iyiliği kriterleri ile Pearson korelasyon katsayısı (*r*), R²_{Adj} ve R² değerleri kullanılmıştır. Her iki sürüde en yüksek R² değeri CART algoritması ile elde edilmiştir. Bu çalışma ile yetiştirci koşullarında yapılan seleksiyon çalışmalarında CART algoritması iyi bir araç olarak değerlendirilebilir.

The Effects of Raising Type on Performances of Some Data Mining Algorithms in Lambs

ABSTRACT

The present study was conducted to compare the predictive performances of CART, CHAID and Exhausted CHAID algorithms to estimate weaning weight (WW) of Awassi lambs raised in elite and base flock. In accordance with this purpose, 331 Awassi lamb (born between November- and January in 2014-2015) records were collected from base flock and elite flock in Toprakkale, Osmaniye province. Birth weight (BW) of Awassi lambs in elite flock and base flock were 4.11 ± 0.07 kg and 4.92 ± 0.05 kg, respectively; weaning weight (WW) of Awassi lambs in elite and base flock were 14.35 ± 0.12 kg and 13.89 ± 0.16 kg, respectively. In both flock (elite and base), dam age at birth, birth type, sex, month, weaning time and birth weight were used as independent variables, while weaning weight was used as dependent variable. In order to compare CART, CHAID and Exhausted CHAID algorithms RMSE, MAPE, RAE, SD_{ratio} and MAD goodness of fit criteria and Pearson correlation coefficient (*r*), R²_{Adj} and R² were used. The highest R² value in both flocks was obtained by the CART algorithms. As a result, in this study, CART algorithm can be evaluated as a good tool in selection studies performed in growing conditions.

To Cite : Tatlıyer A 2020. Kuzularda Yetiştirme Tipinin Regresyon Ağacı Algoritmaları Tahmin Performansına Etkisi. KSÜ Tarım ve Doğa Derg 23 (3): 772-780. DOI: 10.18016/ksutarimdoga.vi.651232

GİRİŞ

Koyunculuk dünyada olduğu gibi Türkiye'de

hayvansal üretimde önemli bir yere sahiptir. Türkiye'de koyun yetiştirciliği ile geçimini sağlayan

insanların sosyo-ekonomik düzeyi ve Türkiye'nin gerek coğrafi yapısından dolayı koyun yetiştiriciliğinin doğal mera alanlarına dayalı olması koyunculuğun minimum maliyetle yapılabilen bir yetiştirmeye faaliyeti olduğunun göstergesi niteliğindedir (Koyuncu, 2012).

TÜİK verilerine göre Türkiye koyun varlığı 2017 yılında 33 milyon başın üzerine çıkmıştır (TUİK, 2018). Bununla birlikte hayvanların genetik niteliği beklenenin oldukça altındadır (Yağcı ve ark., 2018). Yetiştiriciliğin öğrenilen geleneksel üretmeye dayalı olması, yetiştiricilerin örgütlenmemesi, hayvansal ürünlerin pazar fiyatını belirleyen devlet kurumlarının özelleştirilmesi ve küçükbaş yetiştirciliğine yapılan teşviklerin yetersizliği nitelikli damızlık hayvanların yetiştirememesindeki etkili etmenler olduğu söylenebilir.

Küçükbaşa genetik niteliğin belirlenmesi ve çağdaş bir yönelime doğru ilerlenebilmesi amacıyla 2005 yılında Tarımsal Araştırmalar ve Politikalar Genel Müdürlüğü (TAGEM) tarafından başlatılan devlet desteği alt yapısı ile "Halk Elinde Küçükbaş Hayvan İslahi Ülkesel Projesi" uygulanmaya başlanmıştır (Cengiz ve ark., 2015). Söz konusu 2005 yılında devreye sokulup 2006 yılında genişletilen proje ile yetiştiricilerin nitelikli damızlık hayvan yetiştirmesi hususunda bilgi ve deneyimlerini artırarak kayıtlı hayvancılığa geçiş uygulamaya konulmuştur (Yağcı ve ark., 2018). Uygulanan projede numaralandırma ile kayıt sistemi oluşturulmuş, kaydı alınan hayvanlardan oluşan sürüye "elit" alınmayan sürünlere ise "taban sürü" denilmiştir. Halk elinde ıslah projelerinde sürü yetleştirme planında taban sürülerden alınan en iyi dişiler elit sürülere aktarılmıştır. Bu süreçte elit sürüler ile gelecekteki hayvan ıslahı programlarına yön verilmesi hedeflenmiştir.

Söz konusu uygulanan bu girişimle ıslah programlarının hayatı aktarılabilmesi için muazzam büyülükteki veri kayıtları oluşturulmuştur. Bu verilerin yorumlanabilmesi ve anlamlı bir hale getirilebilmesi için çeşitli istatistiksel yöntemler kullanılmaktadır. Ne var ki araştırmacılar tarafından en çok tercih edilen istatistiksel metot Genel Doğrusal Modeller (GLM) olmuştur (Eyduran ve ark., 2008). GLM gibi modellerde verilerin normal dağılması gibi bazı varsayımların yerine getirilmediği durumlarda istatistiksel yöntemler yetersiz kalmakta ve yöntemlerin güvenirliği de düşmektedir (Khan ve ark., 2014). Bununla birlikte kullanılan istatistiksel yöntemlerin etkinliği uygulanacak ıslah programlarına da yön vermektedir. Bilgisayar teknolojisindeki gelişmelere paralel olarak karmaşık yapıdaki verilerin çözümlemesinde kolaylık sağlayan veri madenciliği uygulamalarına yönelik son yıllarda giderek artmış ve hayvancılık alanı da dahil olmak üzere çeşitli alanlarda kullanılmaya başlanmıştır (Temel ve ark., 2010).

Küçükbaş hayvancılık araştırmalarında veri madenciliği yöntemlerinden CART, CHAID ve Exhausted CHAID algoritmalarının kullanıldığı pek çok çalışma bulunmaktadır (Olfaz ve ark., 2019; Balta ve Topal, 2018; Ali ve ark., 2015). Bu çalışmanın amacı CART, CHAID ve Exhausted CHAID algoritmalarının etkinliğini sürü tipine göre (elit sürü ve taban sürü) mukayese ederek hayvan ıslahı projelerindeki temel taş olan elit sürüden elde edilen verilerin yorumlanabilmesini sağlamak, İvesi kuzularında ırk standartlarını belirleyebilmek ve seleksiyonda yardımcı olabilmektir.

MATERIAL ve METOT

Bu çalışmada Osmaniye ili Toprakkale ilçesinde bulunan iki farklı işletmeden elde edilen toplamda 331 baş İvesi ırkı kuzu kayıtları kullanılmıştır. Kuzu kaydı alınan işletmelerin birinin yetleştirme tipi elit (ana-baba kaydı düzenli olarak tutulmuş) diğerinin yetleştirme tipi ise taban sürüdür (ana-baba kaydı düzenli olarak tutulmamış).

Araştırmaya dahil olan işletmelerde koyunlar kiş aylarında ağılda tutulmuş, beslenmelerinde ise ağırlıklı olarak kuru ot samanı, yer fıstığı samanı ve pamuk tohumu küspesi verilmiştir. Bahar döneminde koyunların beslenmesi yaylaya veya meraya dayalı olmuştur ve ek yem verilmemiştir. Her iki sürüde kuzu kayıtları 2014-2015 sezonundan alınmıştır. Elit sürüde doğular Aralık-Ocak aylarında gerçekleşirken, taban sürüde Kasım-Aralık-Ocak aylarında gerçekleşmiştir. Her iki sürüde serbest koç katımı Haziran ayı ortasından itibaren yapılmıştır. Doğumu takiben kuzulara kulak külesi takılmış ve 10 grama duyarlı hassas terazi ile tارتılmıştır. Kuzular ortalama 60 günlük yaşta sütten kesilmiş olup sütten kesim ağırlıkları kayda alınmıştır. Bu çalışmada toplamda 331 kuzunun sütten kesim ağırlığının tahmini CART, CHAID ve Exhausted CHAID algoritmaları kullanılarak tahmin edilmeye çalışılmıştır. Her iki sürüde (elit ve taban) ananın doğumdaki yaşı, doğum tipi, cinsiyet, doğum ayı, sütten kesim zamanı ve doğum ağırlığı bağımsız değişken olarak kullanılırken, sütten kesim ağırlığı bağımlı değişken olarak kullanılmıştır.

Veri madenciliği karmaşık yapıdaki veri setlerinden anlamlı bilgileri kullanarak tahminleme yapabilen bir yöntemdir (Küçükönder ve ark., 2014). Veri madenciliğinde en çok kullanılan yöntem ağaç yapısına dayalı sınıflama ve regresyon ağaçları algoritmalarıdır. Bağımlı değişkenin sürekli olması ile oluşan yapıya "Regresyon Ağacı" denilmektedir (Koç, 2016). Bağımlı değişkenin kategorik olması ile oluşan yapıya ise "Sınıflama Ağacı" olarak tanımlanmaktadır (Oruçoglu, 2011).

Regresyon ağacı metodu parametrik testler için önemli olan varsayımlar gerektirmeyen (normalilik, homojenlik gibi), aynı zamanda çoklu bağlantılılık,

kayıp gözlem ve üç değerlerden etkilenmeyen görsel üstünlüğe sahip bir metottur (Mendeş ve Akkartal, 2009).

Regresyon ağacı metodu ile oluşturulan diyagramla bağımsız değişkenlerin aralarındaki etkileşim ve bağımlı değişkenleri hangi bağımsız değişkenlerin etkilediği kolaylıkla görülmektedir. Aynı zamanda regresyon ağacı metodu, ağaç yapılarının oluşturulmasında kullanılan karar kurallarının yine kolaylıkla anlaşılmasına izin vermektedir (Akşahan ve Keskin, 2015).

Regresyon ağacında en popüler olan algoritmalar ise sürekli, nominal ve ordinal yapıdaki değişkenlerin çözümlenmesinde rahatlıkla kullanılabilen CART (Classification and Regression trees; Sınıflama ve Regresyon Ağacı Algoritması), CHAID (Chi Squared Automatic Interaction Detector; Otomatik Ki-Kare Etkileşim Belirleme) ve Exhausted CHAID (Kapsamlı CHAID) algoritmalarıdır. CART, CHAID ve Exhausted CHAID algoritmalarının her bir aşamada ağaç yapısının oluşturulması, bölünmenin sonlandırılması ya da budama ve her bölümmede her bir bağımsız değişken için bağımlı değişkenin değerini tahmin edilerek en uygun ağaç yapısının elde edilmesi olarak üç ana işlevi bulunmaktadır (Loh, 2011; Moghadam ve ark., 2016; Ali ve ark., 2015; Chang ve ark., 2006).

Ağaç yapısına dayalı bu algoritmaların CART, Breiman ve ark. tarafından (1984) yılında geliştirilmiş olup, her düğümde (node) iki yavru düğüm (child node) oluşturan ve bu işlemi gözlenen değişken ile tahmin değişkeni arasında en güçlü korelasyonun elde edildiği homojen düğümlerin oluşmasına kadar devam ettiren bir algoritmadır (Camdeviren ve ark., 2007). CART algoritması aşırı dallanmaya eğilim gösteren bir algoritma olmasından dolayı, uygulamada budama işleminin araştırmacı tarafından bizzat devreye sokulması gerekmektedir. Bununla birlikte, üzerinde durulan veri setini eğitim ve test olmak üzere setlere ayrılması (genellikle bu oranın 80:20 olarak tavsiye edilir), ağaç derinliğinin hassas bir şekilde tanımlanması ve ebeveyn-yavru düğümdeki birey sayısının ayarlanması gerekmektedir. Aksi takdirde non-parametrik ve görsel olan CART algoritmalarının yorumlanması güç olabilmektedir. Veri setinin eğitim-test olarak ayrılması hususunda diğer bir alternatif ise veri seti üzerine ortalama bir hata değerini hesaplayabilmek için çarpraz geçerlilik değerinin 10 olarak alınmasıdır (Sevgenler, 2019).

CHAID ve Exhausted CHAID algoritmaları ile en uygun ağaç yapısının elde edilmesi üç aşamadan oluşmaktadır: birleştir, böl ve durdur. Kass (1980) tarafından geliştirilen CHAID algoritmasında, CART algoritmasının aksine ikiden fazla bölünme gerçekleşmektedir. Exhausted CHAID algoritması ise sadece birleştirme aşaması farklı olan CHAID algoritmasının geliştirilmiş halidir (Ali ve ark., 2015).

CHAID ve Exhausted CHAID algoritmaları ki-kare analizinde önemli sonuç çıkışına kadar analize devam eder (Wilkinson, 1992). CART algoritmasında ise, bağımlı değişkendeki gözlenen ve beklenen değerler arasında en güçlü korelasyon elde edilinceye analiz devam eder (Camdeviren ve ark., 2007).

Bu üç ağaç yapısına dayalı veri madenciliği algoritmaları arasındaki bir diğer fark; CART algoritmasında bölümme kriteri Gini index kuralına düzeltme yapılrken, CHAID ve Exhausted CHAID algoritmasında Bonferroni kriterine göre düzeltme yapılmaktadır (Hastie ve ark., 2001; Celik ve ark., 2017).

Hayvan yetiştirme ve ıslahı alanında çeşitli araştırmalarda sıkılıkla ağaç yapısına dayalı veri madenciliği algoritmalarına başvurulmaktadır (Balta ve Topal, 2018; Grzesiak ve Zaborski, 2012; Olfaz ve ark., 2019). CART, CHAID ve Exhausted CHAID algoritmalarının performanslarının karşılaştırılabilmesi amacıyla Ali ve ark.'nın (2015) bildirdiği uyum iyiliği kriterleri kullanılmıştır. Bunlar aşağıda belirtildiği gibidir:

Bağımlı değişken bakımından gerçek değerler ve tahmin edilen değerler arasındaki Pearson korelasyon katsayısı (r),

Varyasyon katsayısı (coefficient of variation, coeffvar):

$$\text{coeff var} = \frac{S}{\bar{X}} * 100$$

Determinasyon	Katsayısı	(Coefficient of Determination, R^2)	$R^2 = \left[1 - \frac{\sum_{i=1}^n (y_i - y_{ip})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} \right]$
---------------	-----------	--	---

Düzeltilmiş Coefficient of Determination, R^2_{Adj}	Determinasyon Katsayısı (Adjusted Determination, R^2_{Adj})	$R^2_{Adj} = \left[1 - \frac{\frac{1}{n-k-1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\frac{1}{n-k} \sum_{i=1}^n (Y - \bar{Y})^2} \right]$
---	--	--

Hata kareler ortalamasının karekökü (Root-mean-square error, RMSE) $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{ip})^2}$

Ortalama hata (Mean error, ME) : $ME = \frac{1}{n} \sum_{i=1}^n (y_i - y_{ip})$

Ortalama mutlak sapma (Mean absolute deviation, MAD): $MAD = \frac{1}{n} \sum_{i=1}^n |y_i - y_{ip}|$

Standart sapma oranı (Standard deviation ratio, SD_{ratio}): $SD_{ratio} = \frac{s_m}{s_d}$

Global nispi yaklaşık hata (Global relative

$$\text{approximation error, RAE: } RAE = \sqrt{\frac{\sum_{i=1}^n (y_i - y_{ip})^2}{\sum_{i=1}^n y_i^2}}$$

Ortalama mutlak yüzde hata (Mean absolute

$$\text{percentage error, MAPE: } MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_{ip}}{y_i} \right| \cdot 100$$

n : Veri setindeki hayvan sayısı; k : Modelde bulunan parametrelerinin sayısı; y_i : i. hayvana ait bağımlı değişken değeri; y_{ip} : i. hayvana ait bağımlı değişkenin tahmin değeri; s_m : Model hata terimlerinin standart sapması ve s_d : Bağımlı değişkenin standart sapmasını ifade etmektedir.

En uygun model seçiminde yukarıda adı geçen uyum iyiliği kriterleri kullanılmıştır. Model karşılaştırmaları en düşük RMSE, MAPE, RAE, coeffvar, SD_{ratio} ve MAD değerleri ile en yüksek Pearson korelasyon katsayısı ile birlikte R^{2Adj} ve R²

Cizelge 1. Elit ve taban sürülerde İvesi kuzularının sütten kesim ve doğum ağırlıklarına ait tanımlayıcı istatistikleri

Table 1. Descriptive statistics of weaning weights and birth weights of Awassi lambs in elite and base flock

	Yetiştirme Tipi (Flock type)	N	$\bar{X} \pm S_x$	S	Min	Max
SKA (WW)	Taban (Base)	151	13.88±0.16	2.01	8.77	18.12
	Elit (Elite)	180	14.35±0.12	1.63	10.12	16.71
DA (BW)	Taban (Base)	151	3.91±0.07	0.85	2.05	5.99
	Elit (Elite)	180	4.11±0.05	0.67	2.47	5.14

\bar{X} : Ortalama (Mean); S_x : Standart Hata (Standard Error); S: Standart Sapma (Standard deviation); Min: En az (Minimum) Max: En çok (Maximum)

Elit sürüde İvesi kuzularının sütten kesim ağırlığı 14.35±0.12 iken, taban sürülerde 13.88±0.16 olarak bulunmuştur. Aynı şekilde elit sürüde İvesi kuzularının doğum ağırlığı 4.11±0.05 iken taban sürüde 3.91±0.07 olduğu görülmüştür.

Varyans analizleri uygulamalarında sağlanması gereken "normalilik" gibi çeşitli varsayımlar regresyon

Cizelge 2. Elit ve taban sürülerde sütten kesim ağırlığı (SKA) için regresyon ağaçları algoritmalarının uyum iyiliği kriterleri

Table 2. Goodness of fit criteria of regression tree algorithms for weaning weights in elite and base flock

SKA (WW)	Elit Sürü (Elite) (n=180)				Taban Sürü (Base) (n=151)		
	Uyum iyiliği kriterleri	CHAID	Exhausted CHAID	CART	CHAID	Exhausted CHAID	CART
<i>pearcorr</i>	0.991	0.989	0.997		0.781	0.781	0.814
<i>Coeffvar</i>	1.157	1.656	0.945		9.046	9.047	8.407
<i>Sdratio</i>	0.134	0.146	0.083		0.624	0.625	0.580
<i>RMSE</i>	0.217	0.237	0.135		1.250	1.251	1.163
<i>ME</i>	-0.001	-0.001	0.001		0.002	0.002	-0.001
<i>RAE</i>	0.015	0.016	0.009		0.088	0.089	0.083
<i>MAPE</i>	1.174	1.257	0.756		7.542	7.543	6.907
<i>MAD</i>	0.155	0.165	0.102		0.991	0.991	0.917
<i>Rsq</i>	0.982	0.979	0.993		0.610	0.610	0.663
<i>AdjRsq</i>	0.982	0.979	0.993		0.602	0.602	0.656

değerine göre yapılmıştır.

IBM SPSS 23 programı (Analysis > Classify > Tree) ile ağaç yapısına dayalı CART, CHAID ve Exhaustive CHAID algoritmaları analizleri yapılmıştır. En yüksek tahmin performansını elde etmek amacıyla ebeveyn-yavru düğümdeki minimum hayvan sayıları elit sürü ve taban sürüde 12:6 olarak ayarlanmıştır. Çapraz geçerlilik değeri ise her iki sürüde 10 olarak ayarlanmıştır. Bu analizlerin sonucunda bulunan bağımlı değişkene ait tahmini değerler ile gözlenen (gerçek) değerler bir not defteri dosyasına kaydedilmiş ve ardından ücretsiz olarak indirilip kullanılabilen R STUDIO yazılımı ile uyum iyiliği kriterleri hesaplanmıştır.

BULGULAR ve TARTIŞMA

Aynı bölgede bulunan ve yetiştirme tipi elit ve taban olan iki farklı işletmede bulunan İvesiirk kuzuların sütten kesim ağırlıklarına ait tanımlayıcı istatistikleri Cizelge 1'de özetlenmiştir.

ağacı analizlerinde gerekli değildir ve veriler görsel olarak kolaylıkla yorumlanabilmektedir (Celik ve ark., 2017). Çapraz geçerlilik değerinin 10 olarak ayarlandığı bu çalışmada regresyon ağaçları (CART, CHAID ve Exhaustive CHAID) algoritmalarının uyum iyiliği kriterlerine göre sonuçları Cizelge 2'de belirtildiği gibidir.

Çizelge 2'de özetlendiği gibi, elit sürüde sütten kesim ağırlığının gözlenen değerleri ile regresyon ağacı algoritmaları (CHAID, Exhaustive CHAID ve CART) kullanılarak tahmin edilen değerleri arasındaki Pearson korelasyon katsayıları sırasıyla 0.991, 0.989 ve 0.997'dir ve oldukça yüksektir. Taban sürüde Pearson korelasyon katsayıları daha düşüktür ve tahmin edilen değerlerle gözlenen değerler arasındaki Pearson korelasyon katsayıları sırasıyla 0.781, 0.781 ve 0.814'tir. Farklı büyülükteki verileri standardize ederek mukayese edebilen bir değişim ölçüsü olan varyasyon katsayılarına bakıldığında, elit sürüdeki varyasyon katsayılarının taban sürüde elde edilen değerlerden oldukça farklı olduğu görülmüştür. Elit sürüde varyasyon katsayısı (*coeffvar*) 0.9 ile 1.6 arasında değişkenlik göstermiştir, taban sürüde VK ise 8.4 ile 9.04 arasında değişmiştir. Elite sürüdeki regresyon ağacı algoritmalarının *coeffvar* sıralaması Exhaustive CHAID > CHAID > CART şeklindedir. Taban sürüdeki regresyon ağacı algoritmalarının *coeffvar* sıralaması ise Exhaustive CHAID = CHAID > CART şeklindedir. Standart sapma oranı (*SD_{ratio}*) elit sürüde 0.08-0.14 arasında değişirken, taban sürüde daha yüksek olup 0.58 ile 0.62 arasındadır.

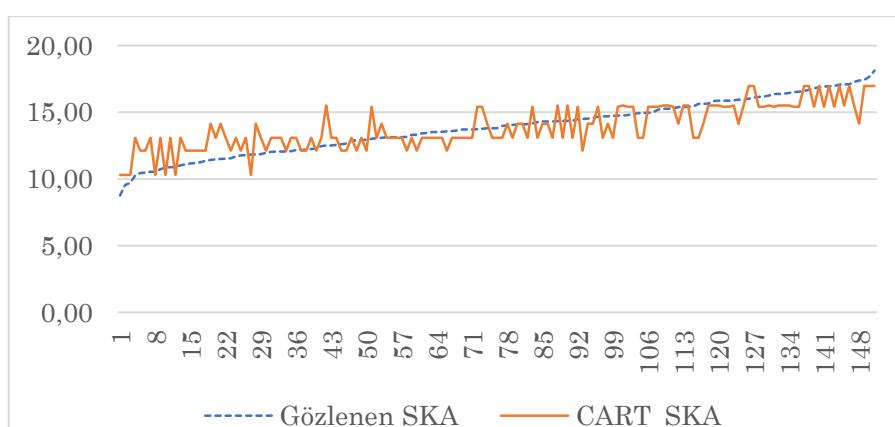
Çizelge 1'de görüldüğü üzere, taban sürüdeki regresyon ağaçlarına ilişkin uyum iyiliği kriterlerine bakıldığında CHAID ve Exhausted CHAID algoritmalarına ait değerlerin hemen hemen aynı olduğu görülmüştür. Bununla birlikte taban sürüde determinasyon katsayıları ($R^2=0.61$) ve düzeltilmiş determinasyon katsayıları ($R^2_{Adj}=0.60$) her iki algoritma için aynı bulunmuştur. Taban sürüde CART algoritması ile elde edilen determinasyon (R^2) ve düzeltmiş determinasyon katsayısı (R^2_{Adj}) CHAID ve Exhausted CHAID algoritmaları ile elde edilen değerlerden daha yüksek olup sırasıyla 0.66 ve 0.65 olduğu görülmüştür. Elite sürüde ise CHAID, Exhausted CHAID ve CART ile elde edilen R^2 ve R^2_{Adj} değeri taban sürüdeki değerlerden yüksek olup 0.98 ile 0.99 arasında değişmiştir.

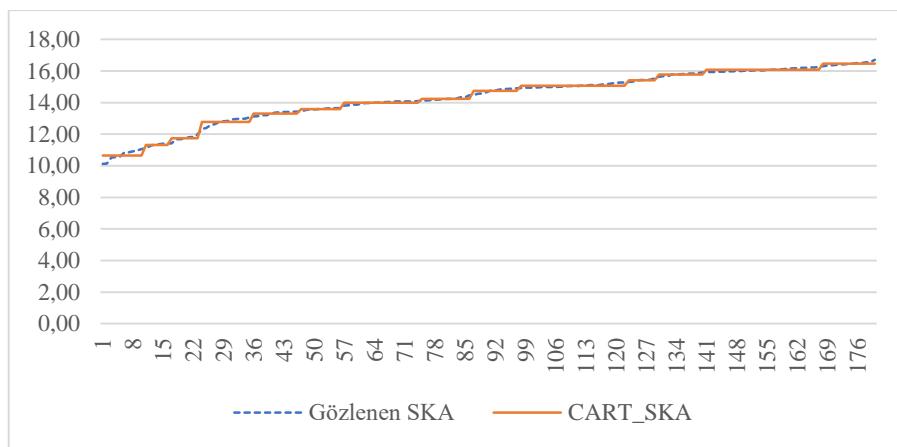
Model karşılaştırmaları en düşük *RMSE*, *MAPE*, *RAE*, *coeffvar*, *SD_{ratio}* ve *MAD* değerleri ile en yüksek Pearson korelasyon katsayına, R^2_{Adj} ve R^2 değerine göre değerlendirilmiş ve buna göre hem elit sürüde hem de taban sürüde CART algoritmasının en uygun algoritma olduğu görülmüştür. R^2 değerleri bakımından değerlendirildiğinde, elit sürüdeki sütten kesim ağırlığındaki varyasyonun % 99.3'ü, taban sürüdeki sütten kesim ağırlığındaki varyasyonun % 66.3'ü CART algoritması ile önemli bulunan bağımsız değişkenler tarafından açıklanmıştır.

Bununla birlikte CART algoritmasının tahmin performansının hem taban sürüde hem de elit sürüde gerçek ve tahmin değerler arasındaki uyumu gösteren grafikler Şekil 1 ve Şekil 2'de belirtilmiştir. Buna göre CART algoritması elit sürüde hayvanların sütten kesim ağırlığına ilişkin tahmin değerlerinin gerçek değerleri daha iyi yansittığı görülmektedir.

Elite sürüdeki CART algoritması ile elde edilen ağaç diyagramı Şekil 3'de görüldüğü gibidir. Şekil 3 incelediğinde CART algoritması ile modelde bulunan tüm bağımsız değişkenler elenip (doğum ağırlığı, doğum tipi, cinsiyet, doğum zamanında ananın yaşı, sütten kesim zamanı) sadece en anlamlı bağımsız değişkenin doğum ağırlığı olduğu görülmektedir. Elite sürüde üç ağaç derinliği olup dallanmalar sadece doğum ağırlığı faktörüne göre form almıştır. Düğüm 0 diğer bir ifade ile kök düğümelit sürüdeki kuzuların sütten kesim ağırlığına ilişkin bazı tanımlayıcı istatistiklerini de kısaca özetlemektedir (14.35 ± 1.63).

Birinci ağaç derinliğinde bulunan Düğüm 1'de ve ikinci ağaç derinliğinde bulunan Düğüm 4'de bulunan İvesi ırkı kuzular kendi içinde homojen olduklarından dolayı başka ağaç derinliklerinde düğümlere bölünmemiştir. Bununla birlikte kendi içinde homojen olup dallanma yapmayan düğümlere "Terminal Düğüm" denilmektedir (Eydurhan ve ark., 2008). Şekil 3 incelediğinde, Düğüm 5 ve Düğüm 6 da Terminal düğümler olarak tanımlanabilmektedir.

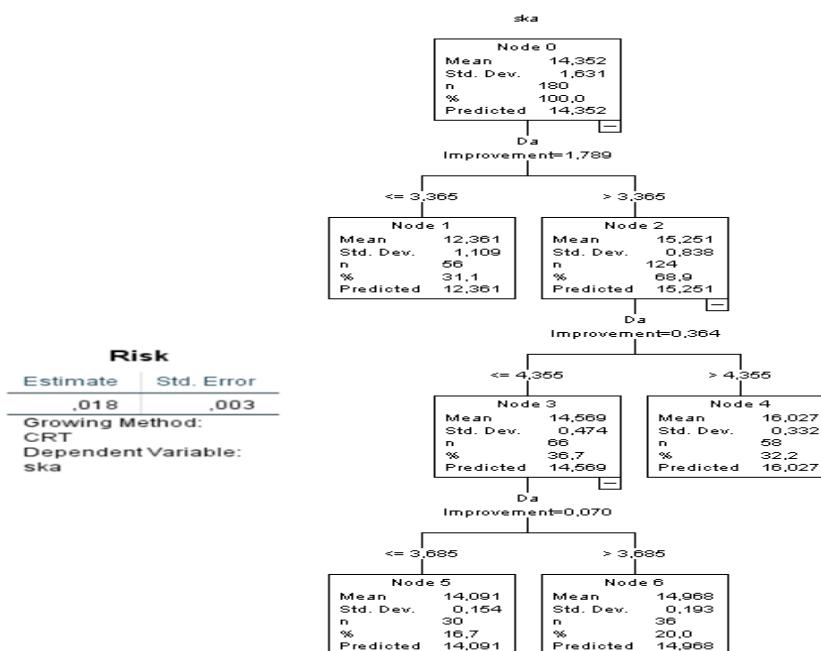




Şekil 2. Elit sürüde uygulanan CART algoritması için uyum grafiği
 Figure 2. Graph of goodness for CART algorithms in base flock

Elit sürüde en yüksek sütten kesim ağırlığı Da > 4.35 olan kuzularda; en düşük sütten kesim ağırlığı ise Da ≤ 3.36 olan kuzularda görülmüştür. CART algoritması ile sadece doğum ağırlığına bakarak sütten kesim

ağırlığı hakkında yorum yapılabılır; en yüksek ve en düşük sütten kesim ağırlığına sahip olan kuzular görsel olarak kolaylıkla ayırt edilebilmektedir.



Şekil 3. Elit sürüdeki CART algoritması ile elde edilen regresyon ağaçları.
 Figure 3. Regression tree diagram obtained by CART algorithm in elite flock

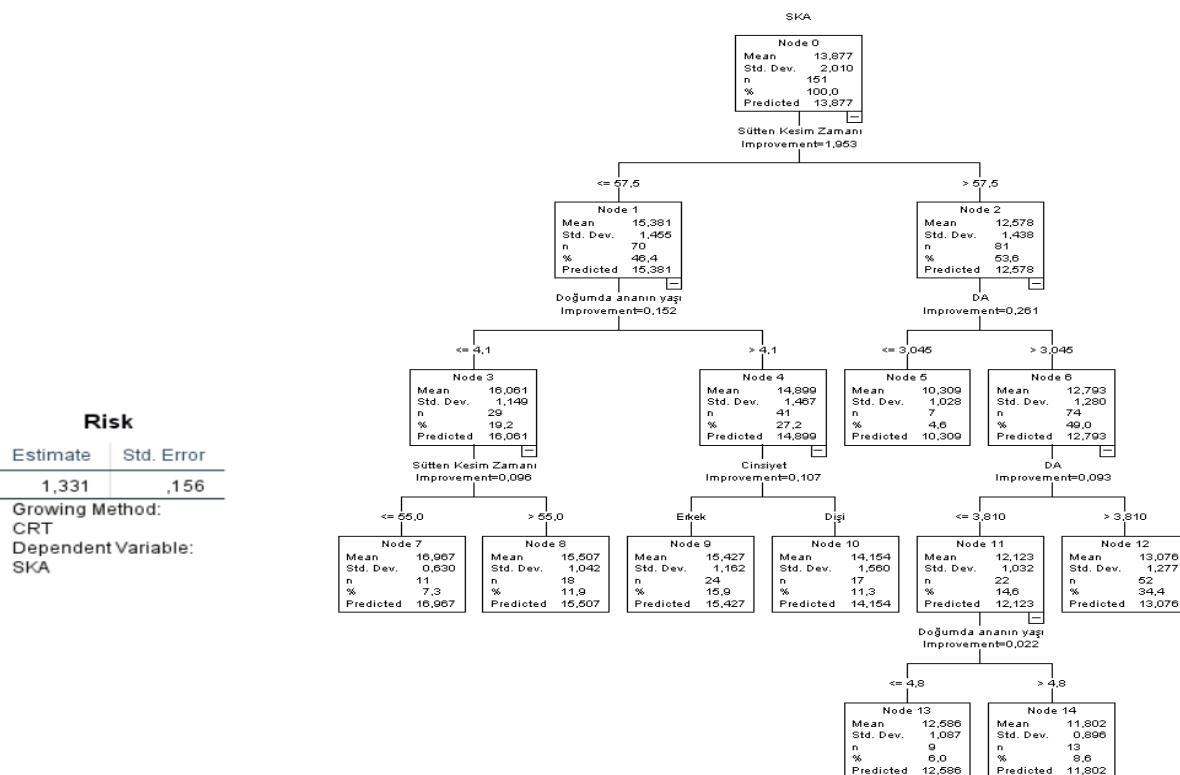
Çizelge 2'de görüldüğü gibi, uyum iyiliği kriterlerine bakıldığından (en düşük RMSE, MAPE, RAE, coeffvar, SD_{ratio} ve MAD değerleri ile en yüksek Pearson korelasyon katsayıısı ile birlikte R²_{Adj} ve R² değerleri) elit sürüde olduğu gibi taban sürüde de CART algoritması sütten kesim ağırlığının tahminlenmesinde en iyi performansı göstermiştir. Elde edilen sonuçlara göre, elit sürüde gözlenen değerler ile tahmin edilen değerler arasındaki en yüksek korelasyon (0.814), en yüksek belirleme

katsayıısı ($R^2=0.663$) ve en düşük standart sapma oranı (0.580) CART algoritmasından elde edilmiştir.

Taban sürüdeki CART algoritması ile elde edilen ağaç diyagramı Şekil 4'deki gibidir. Taban sürüdeki CART algoritması ile şekillenen regresyon ağaçları diyagramı elit sürüdeki ağaç diyagramına göre daha fazla dallanma göstermiştir. Şekil 4'de taban sürüde daha fazla faktörün sütten kesim ağırlığını etkilediği görülmektedir. CART algoritması ile şekillenen diyagramda (Şekil 4), sütten kesim ağırlığını etkileyen

değişkenler doğum ağırlığı faktörüne ilaveten sütten kesim zamanı, cinsiyet ve ananın doğumdaki yaşıdır. Doğum tipi ise sütten kesim ağırlığında etkili bir değişken olmamıştır. Taban sürüde CART algoritması ile oluşan regresyon ağacı dört ağaç derinliğinden oluşmaktadır. Düğüm 0, bu araştırmaya konu olan tüm kuzuların ihtiiva ettiği heterojen gruptur. Düğüm 0, sütten kesim zamanına göre Düğüm 1 (70 baş İvesi kuzusunun bulunduğu alt grup, ortalama SKA=15.381 kg) ve Düğüm 2 (81 baş İvesi kuzusunun olduğu alt grup, ortalama SKA= 12.578 kg) olmak üzere ikinci ağaç derinliğindeki Düğüm 1, ananın doğumdaki yaşı değişkenine göre Düğüm 3 (ananın doğumdaki yaşı ≤ 4.1 olan kuzular 16.061 ± 1.149 kg; n=29) ve Düğüm 4 (ananın doğumdaki yaşı > 4.1 olan kuzular 14.899 ± 1.467 kg; n=41) olmak üzere iki alt gruba ayrılmıştır. Düğüm 3, ikinci ağaç derinliğindedir ve sütten kesim zamanı değişkenine göre Düğüm 7 (11 baş İvesi kuzusu ve ortalaması 16.967 ± 0.630 kg) ve Düğüm 8 (18 baş İvesi kuzusu ve ortalaması 15.507 ± 1.042 kg) olmak üzere iki gruba ayrılmıştır. Sütten kesim zamanı 57.5 gün ve altında olan kuzular arasında en yüksek sütten kesim ağırlığına (16.967 kg) sahip olan kuzular Düğüm 7'de görülmektedir. Düğüm 7 ve 8'de, kendi içinde homojen olduklarından dolayı tekrar bölünme gerçekleşmemiştir. İkinci ağaç derinliğindeki Düğüm 4, cinsiyet faktörü bakımından Düğüm 9 ve 10

olmak üzere iki alt gruba bölünmüştür. Yirmidört baş İvesi kuzusunun oluşturduğu Düğüm 9'da ortalama 14.154 ± 1.560 kg canlı ağırlığa sahip erkek kuzular bulunmaktadır iken, Düğüm 10'da ortalama 14.154 ± 1.560 kg canlı ağırlığa sahip dişi kuzular bulunmaktadır. Cinsiyet etkisi bakımından bölünmenin düşümlerde erkek kuzuların bulunduğu Düğüm 9, dişi kuzuların bulunduğu Düğüm 10'dan daha ağırdır ki bu sonuç da beklenen bir durumdur. Sütten kesim zamanı 57.5'den fazla olan kuzuların bulunduğu grup (Düğüm 2), doğum ağırlığı özelliği bakımından ikinci ve üçüncü ağaç derinliklerinde tekrarlamalı iki alt gruba ayrılmıştır. Üçüncü ağaç derinliğinde doğum ağırlığı 3.810 kg ve altında olan kuzuların olduğu Düğüm 11 (22 baş İvesi kuzusu ve ortalaması 12.123 ± 1.032 kg) ananın doğumdaki yaşı bakımından Düğüm 13 (12.586 ± 1.087 kg; n=9) ve Düğüm 14 (11.802 ± 0.896 kg; n=13) olmak üzere tekrar iki alt gruba bölünmüştür. Taban sürüde CART algoritması ile elde edilen diyagramda, Düğüm 5, 7, 8, 9, 10, 12, 13 ve 14, kendi içinde homojen oldukları terminal düşümlerdir ve bu düşümlerden sonra tekrar bölünme gerçekleşmemiştir. Sütten kesim zamanı 57.5'in üzerinde olan kuzular içerisinde en yüksek sütten kesim ağırlığına sahip olan kuzular, doğum ağırlığı 3.810'dan fazla olan kuzuların bulunduğu Düğüm 12'dedir (13.076 ± 1.277 kg; n=52). Bu sonuctan, doğum ağırlığının sütten kesim ağırlığına pozitif etkilediği anlaşılmaktadır.



Şekil 4. Taban sürüdeki CART algoritması ile elde edilen regresyon ağacı diyagramı.

Figure 4. Regression tree diagram obtained by CART algorithm in base flock

Bu çalışma, elit sürüdeki analardan doğan kuzuların doğum ağırlıkları ve sütten kesim ağırlıkları taban sürüdeki analardan doğan kuzulardan daha yüksektir ve doğum ağırlıkları ile sütten kesim ağırlıkları arasındaki fark istatistiksel olarak önemli olduğunu göstermiştir ($P<0.05$). Sezgin ve ark. (2012), doğum ağırlığı özelliğinin sürü tipi bakımından kuzular arasındaki farkı istatistiksel olarak anlamlı bulurken, Balta ve Topal (2018) yaptıkları çalışmada elit sürüde doğan kuzuların doğum ağırlığının taban sürüde doğan kuzulardan daha yüksek olduğunu bildirmiştir ancak bu farklılığın istatistiksel olarak önemli olmadığını belirtmişlerdir.

Mevcut çalışmada regresyon ağacı diyagramı bakımından elit ve taban sürüde sütten kesim ağırlığını etkileyen faktörler değişmektedir. Buna göre, elit sürüde yetişmiş olan kuzuların sütten kesim ağırlığının tahminlenmesinde en anlamlı ve hatta tek anlamlı bağımsız değişkenin doğum ağırlığı olduğu görülmüşken, taban sürüde yetişmiş olan kuzuların sütten kesim ağırlığının tahminlenmesinde önemli bağımsız değişkenler doğumdaki ananın yaşı, yavrunun cinsiyeti, sütten kesim zamanı ve doğum ağırlığı değişkenleridir. Doğum ayı ve doğum tipi her iki sürüde sütten kesim ağırlığının tahminlenmesinde etkili olmamıştır. CART algoritması ile elde edilen regresyon ağacı diyagramında elit sürüde risk değeri (0.018) ve risk değerinin standart hatası (0.003) oldukça düşüktür. Taban sürüde ise, risk değeri (1.331) ve risk değerinin standart hatası (0.156) elit sürüden daha yüksektir. Ancak her iki sürüde de determinasyon katsayısı (R^2) bakımından CART algoritması iyi bir performans sergilemiştir. Yine çalışmada risk değeri ve risk değerinin standart hatasının düşük olması, sütten kesim ağırlığının açıklanmasında CART algoritmasının yeterli olabileceğini göstermektedir. Balta ve Topal (2018), yaptıkları çalışmada CART algoritmasının performansını doğum ağırlığının tahminlenmesinde incelemiştir. Araştırmada, bulunan R^2 değeri (0.862), mevcut çalışmadaki taban sürüdeki R^2 değerinden (0.663) yüksek, elit sürüdeki R^2 değerinden (0.993) düşüktür. Çelik ve ark (2017), Mengali koçlarında canlı ağırlığın tahminlenmesinde çeşitli veri madenciliği algoritmalarının (CHAID, Exhausted CHAID, CART, MARS, MLP, RBF) performanslarını mukayese etmişlerdir. Araştırmada, en iyi performansı ($R^2=0.920$) CART algoritması ile elde edilmiş ve bu çalışmadaki elit sürüdeki performans değeri ($R^2=0.993$) benzerdir. Mendeş ve Akkartal (2009) CART algoritmasını kullanarak yaptıkları regresyon ağacı analizinde R^2 değeri 0.810 olup, bu çalışmadaki taban sürüde görülen R^2 değerinden yüksek ve mevcut elit sürüde görülen R^2 değerinden düşüktür. Olfaz ve ark. (2018), Karayaka kuzusunun sütten kesim ağırlığının tahminlenmesinde CART algoritması kullanılmışlardır. Çalışmalarında CART

algoritması Karayaka kuzusunun sütten kesim ağırlığındaki varyasyonun % 88'i kullanılan bağımsız değişkenler (yaş, cinsiyet, doğum tipi, doğum ağırlığı ve ölçüm zamanı) tarafından açıklanmıştır ($R^2=0.880$). Yakubu (2012), Uda koyunları üzerinde yaptığı çalışmada bazı morfolojik özellikler ile canlı ağırlığın tahminlenmesinde CART algoritması kullanmışlar ve R^2 değerini 0.618 olarak bildirmiştir. Araştırmacının elde ettiği bu değer mevcut çalışmadaki taban sürüsünde elde edilen değer (0.663) ile benzerdir, ancak mevcut çalışmadaki elit sürüdeki değerden (0.993) düşüktür.

Sonuç olarak, regresyon ağacı analiz yöntemindeki algoritmalar ile yapılan tahminler hayvansal üretimde etkili olabilir. Ebeveyn: yavru düğümleri için tercih edilen minimum hayvan sayıları araştırmadan araştırmaya farklılık göstermekle birlikte minimum hayvan sayısı değeri azaldıkça tahmin performanslarının arttığı görülmüştür. Aynı zamanda üzerinde çalışılan sürülerde tutulan kayıtların doğruluğu, buna benzer çalışmaların etkinliğini doğrudan etkilemektedir. Bu çalışmada elit ve taban sürülerde görülen performans farklılığı bu görüşü desteklemektedir. Dolayısıyla gelecekte yapılması planlanan hayvan ıslahı çalışmalarında, uygulanacak regresyon ağacı algoritmalarında ölçüm hassasiyetinin önemli bir kriter olduğu görülmüştür. Hayvan ıslahı çalışmalarında yapılan ilk adım kayıt tutmaktadır. Dolayısıyla seleksiyona teşkil edecek kayıtlar ile gelecek generasyonların ebeveynlerinin belirlenmesinde elit sürülerin sayısı artırılabilir, yetişiricilere ölçümülerin titizlikle yapılması konusundaki özverinin önemi vurgulanabilir ve tutulan kayıtlarda yeni özellikler eklenebilir. Daha büyük hayvan sürüleri üzerinde çalışmak, kullanılan istatistiksel yöntemlerindeki yanılmaları düşürebilir. Bununla birlikte daha etkin bağımsız değişkenlerin seçilmesi regresyon ağacı yöntemlerinin tahminleme gücünü artırabilir.

Teşekkür

Kayıtlarını kullanmamıza izin vererek bu çalışmanın gerçekleşmesine katkıda bulunan Osmaniye İli Damızlık Koyun Keçi Yetiştiricileri Birliği'ne ve Proje Teknik Elemanlarına teşekkür ederim.

Çıkar Çatışması Beyanı

Makale yazarı herhangi bir çıkar çatışması olmadığını beyan eder.

KAYNAKLAR

- Aksahan R, Keskin İ 2015. Determination of the Some Body Measurements Effecting Fattening Final Live Weight of Cattle by the Regression Tree Analysis. Selçuk Tarım Bilimleri Dergisi, 2(1): 53-59.
Ali M, Eyduran E, Tariq MM, Tirink C, Abbas F, Bajwa

- MA, Baloch MH, Nizamani AH, Waheed A, Awan MA, Shah SH 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(6): 1579–1585.
- Balta B, Topal M 2018. Regression tree approach for assessing the effects of non-genetic factors on birth weight of Hemşin lamb. *Alinteri Zirai Bilimler Dergisi*, 33(1): 65–73.
- Breiman L, Friedman J, Olshen R, Stone C 1984. Classification and regression trees. Wadsworth Int. Group, 37(15): 237-251.
- Camdeviren HA, Yazici AC, Akkus Z, Bugdayci R and Sungur MA 2007. Comparison of logistic regression model and classification tree: An application to postpartum depression data. *Expert Systems with Applications*, 32(4):987-994.
- Celik S, Eydurhan E, Karadas K, Tariq MM 2017. Comparison of predictive performance of data mining algorithms in predicting body weight in Mengali rams of Pakistan. *Revista Brasileira Zootecnia*, 46(11): 863–872.
- Cengiz F, Karaca S, Kor A, Ertuğrul M, Arık İZ, Gökdal, Ö 2015. Küçükbaş hayvan yetişiriciliğinde değişimler ve yeni arayışlar. *Türkiye Ziraat Mühendisliği VIII. Teknik Kongresi* (12-16 Ocak 2015), Ankara.
- Chang LY, Wang HW 2006. Analysis of traffic injury: An application of non-parametric classification tree techniques. *Accident Analysis Prevention*, 38, 1019-1027.
- 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.
- Grzesiak W, Zaborski D 2012. Examples of the use of data mining methods in animal breeding. In: Data mining applications in engineering and medicine (ed. A Karahoca). InTech, Rijeka, Croatia,. in IntechOpen 303–324. DOI: <http://dx.doi.org/10.5772/57353>
- Hastie,T. Tibshirani,R., Friedman J 2001. The Elements of Statistical Learning. Springer, New-York
- Kass, G. 1980. An exploratory technique for investigating large quantities of categorical Data. *Applied Statistics*, 29:2, 119-127.
- Khan MA, Tariq MM, Eydurhan E, Tatliyer A, Rafeeq M, Abbas F, Rashid N, Awan MA, Javed K 2014. Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. *Journal of Animal and Plant Science*, 24(1):120–126
- Koc Y 2016. Application of regression tree method for different data from animal science. MSc Thesis, Igdir University, 58.
- Koyuncu M 2012. Türkiye hayvancılığında küçükbaş hayvancılığının yeri. *Bursa Tarım Kongresi*, (27-29 Eylül 2012), Bursa.
- Küçükönder H, Üçkardeş F, Nariç D 2014. Hayvancılık Alanında Bir Veri Madenciliği Uygulaması: Japon Bildircini Yumurtalarında Döllülüğe Etki Eden Bazı Faktörlerin Belirlenmesi. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 20 (6): 900-908.
- Loh WY 2011. Classification and regression trees. *Wires Data Min Knowl*, 1, 14-23.
- Mendeş M, Akkartal E 2009. Regression tree analysis for predicting slaughter weight in broilers. *Italian Journal of Animal Science*, 8(4):615-624.
- Moghadam MPA, Pahlavani P, Naseralavi, S 2016. Prediction of car following behavior based on the instantaneous reaction time using an ANFIS-CART based model. *International Journal of Transportation Engineering* 4 (2): 109-126.
- Olfaz M, Tırnak C, Önder H 2019. Use of CART and CHAID algorithms in Karayaka sheep breeding. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 25(1): 105–110. DOI: 10.9775/kvfd.2018.20388
- Oruçoglu O 2011. Holstein ırkı ineklerin 305 günlük süt verimini etkileyen çevre faktörlerinin regresyon ağacı ile belirlenmesi. Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Ana Bilim Dalı, Yüksek lisans tezi, , 61 s.
- Sevgenler, H 2019. Keçilere ait kimi özelliklerin canlı ağırlık üzerindeki etkilerini belirlemek amacıyla kullanılan veri madenciliği algoritmalarının (Cart, Chaid ve Mars) karşılaştırılması. İğdır Üniversitesi, Fen bilimleri Enstitüsü, Zootekni Ana Bilim Dah, Yüksek Lisans Tezi, 70 s.
- Sezgin E, Kopuzlu S, Yuksel S, Esenbuga N, Bilgin ÖC 2012. Determination of growth traits and heritabilities of growth characteristics of Hemşin sheep reared in Artvin. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 18: 899-905.
- Temel GO, Ankaralı H, Yazıcı AC 2010. Regresyon Modellerine Alternatif Bir Yaklaşım:MARS. *Türkiye Klinikleri Biyoistatistik*, 2(2): 58-66.
- TUİK 2018: Türkiye İstatistik Kurumu. Erişim: <https://biruni.tuik.gov.tr/hayvancilikapp/hayvancilik.zul>.
- Wilkinson L 1992. Tree Structured Data Analysis: AID, CHAID and CART. In: 1992 Sun Valley, ID, Sawtooth/SYSTAT Joint Software Conference 1–10.
- Yağcı S, Baş S, Tatlıyer A 2018. Şavak Akkaraman Kuzuların Yetiştirici Koşullarında Büyüme ve Yaşama Gücü Özellikleri. *Lalahan Hayvancılık Araştırma Enstitüsü Dergisi* 58 (2): 81-88.
- Yakubu A 2012. Application of Regression Tree Methodology in Predicting the Body Weight of Uda Sheep. *Animal Science and Biotechnologies*, 45 (2): 484-490.