



Selçuk Tarım Bilimleri Dergisi

Sığırlarda Besi Sonu Canlı Ağırlığını Etkileyen Bazı Vücut Ölçülerinin Regresyon Ağacı Yöntemi ile Belirlenmesi

Rifat Akşahan¹, İsmail Keskin^{2*}

¹Dumlupınar İlçe Tarım, Gıda ve Hayvancılık Bakanlığı, Dumlupınar, Kütahya

²Selçuk Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Konya

MAKALE BİLGİSİ

Makale Geçmişi:

Geliş tarihi 10 Ocak 2015

Kabul tarihi 15 Nisan 2015

Anahtar Kelimeler:

CHAID algoritması

Regresyon Ağacı

Siyah Alaca

Simental

Esmer Sığır

ÖZET

Bu çalışmanın amacı, Türkiye'nin Afyon ili Bolvadin ilçesinde yetiştirilen 103 baş (Siyah Alaca, Simental, Esmer ve Melez) tosunun besi sonu canlı ağırlığını (BSCA) etkileyen vücut ölçülerini belirlemektir. Bu amaç için, CHAID algoritmasını esas alan regresyon ağacı metodu kullanılmıştır. Bağımlı değişken olarak, BSCA kullanılmıştır. Bu çalışmada kullanılan bağımsız değişkenler sırasıyla, cidago yüksekliği (CY), sırt yüksekliği (SIY), sağrı yüksekliği (SAY), kuyruk sokumu yüksekliği (KSY), vücut uzunluğu (VU), arka sağrı genişliği (ASG), göğüs derinliği (GD), ve göğüs çevresi (GÇ) ve besi süresi (BS) olmuştur. Bu çalışmada en iyi model kriterlerini elde etmek amacıyla CHAID algoritması için ebeveyn ve çocuk düğümde bulunan minimum tosun sayıları 8:4 olarak ayarlanmıştır. CHAID algoritmasının etkinliğini belirlemek için model kalite kriterleri olarak belirleme katsayısı, düzeltilmiş belirleme katsayısı ve gerçek ve tahmin edilen BSCA değerleri arasındaki Pearson korelasyon katsayısı kullanılmıştır. BSCA etkileyen bağımsız değişkenler; GÇ (Düz. $P < 0.01$), VU (Düz. $P < 0.05$), KSY (Düz. $P < 0.01$) ve BS (Düz. $P < 0.01$) olmuştur. Belirleme katsayısı (%), düzeltilmiş belirleme katsayısı (%) ve gerçek ve tahmin edilen BSCA değerleri arasındaki Pearson korelasyon katsayısı sırasıyla, 87.82 (%), 87.32 (%) ve 0.937 bulunmuştur. Model kalite kriterleri, CHAID algoritması kullanımının yüksek bir tahmin doğruluğu sağladığı göstermiştir. Ortalama olarak, en ağır BSCA (545.850 kg); göğüs çevresi $GÇ > 190$ cm olan tosunların oluşturduğu alt gruptan elde edilmiştir. İkinci olarak en ağır BSCA (505.562 kg); $183 < GÇ < 190$ cm ve $BS > 4$ ay olan tosunların oluşturduğu alt gruptan elde edilmiştir. Sonuç olarak, farklı vücut ölçülerinden BSCA tahmin etmek amacıyla CHAID algoritması kullanımı, hayvan ıslahı çalışmaları için tavsiye edilebilir.

Determination of the Some Body Measurements Effecting Fattening Final Live Weight of Cattle by the Regression Tree Analysis

ARTICLE INFO

Article history:

Received 10 January 2015

Accepted 15 April 2015

Keywords:

CHAID algorithm

Regression tree

Holstein

Simental

Brown swiss

ABSTRACT

The aim of this study is to determine body measurements affecting final live weight (FLW) at fattening period of 103 young (Holstein, Simental, Brown Swiss and crossbreed) bulls reared in Bolvadin district of Afyon province of Turkey. For this aim, a regression tree method on the basis of CHAID (Chi-Squared Automatic Interaction Detector) algorithm were used. FLW was considered as response variable. The independent variables included in the study were withers height (WH), back height (BH), front rump height (FRH), back rump height (BRH), body length (BL), back rump width (BRW), chest dept (CD), chest circumference (CC), and fattening period (FP). Minimum numbers of young bulls in parent and child nodes for CHAID algorithm were assigned as: 8:4, in order to obtain the best model quality criteria for the data evaluated under the study. Model quality criteria like coefficient of determination (%), adjusted coefficient of determination (%), and the Pearson correlation between actual and

* Sorumlu yazar email: ikeskin@selcuk.edu.tr

predicted fattening FLW values were estimated to determine the effectiveness of CHAID algorithm. The independent variables influencing FLW were CC (Adj. P<0.01), BL (Adj. P<0.05), BRH (Adj. P<0.01), and FP (Adj. P<0.01). Coefficient of determination (%), adjusted coefficient of determination (%), and the Pearson correlation between actual and predicted fattening final weight values were found 87.82 (%), 87.32 (%), and 0.937, illustrating that use of CHAID algorithm visually provided a high predictive accuracy. The heaviest FLW (545.850 kg) at fattening period were obtained from the sub-group of the young bulls with CC > 190 cm, averagely. The secondly heaviest sub group of the young bulls with 183 < CC < 190 cm and FP > 4 months provided a FLW of 505.562 kg. Consequently, the use of CHAID algorithm for predicting FLW from various body measurements could be advised to be useful for breeding purposes.

1. Giriş

Bağımsız değişkenin (ya da değişkenlerin) kendi ölçü birimi cinsinden bir birim değişmesine karşılık bağımlı değişkenin (ya da değişkenlerin) kendi ölçü birimi cinsinden ortalama olarak ne kadar değişeceğini gösteren katsayıya regresyon katsayısı denmektedir. Bağımsız değişken veya değişkenler ile bağımlı değişken arasındaki ilişkiyi inceleyen regresyon analizi; basit veya çoklu olabildiği gibi, doğrusallık, normallik, homojenlik, toplanabilirlik gibi bir takım varsayımların yerine getirilmesinden sonra uygulanabilmektedir. Bu varsayımların yerine gelmemesi durumunda veri setindeki orijinal değerler ya logaritmik dönüşüm, ya da karekök dönüşüm yöntemleri gibi bir takım dönüştürme işlemlerine tabi tutularak uygun hale getirilmeye çalışılmaktadır (Efe ve ark., 2000). Parametrik yöntemlerde her ne kadar dönüştürme metodları kullanılarak varsayımlar yerine getirilmeye çalışılırsa da, yapılan analiz neticesinde veri setine ilişkin yanlış sonuçlar elde edilebilmektedir. Bu nedenle alternatifin olmadığı durumlarda verileri dönüştürme yoluna gidilmesi istatistik açıdan daha doğrudur (Kayri ve Boysan, 2008).

Karar ağaçları, bir problemi oluşturan veri setlerinin yapısına göre bir ağaç yapısı şeklinde sınıflandırma ve regresyon modelleri oluşturmaktadır. Söz konusu ağaç yapılarının oluşturulmasında kullanılan karar kurallarının anlaşılabilir olması yöntemin kullanımını yaygın hale getirmiştir. Karar ağaçları, sınıflama ve regresyon probleminin çözümünde çok aşamalı ve ardışık bir yaklaşım ile karmaşık yapıdaki verileri aşamalı bir hale dönüştürerek basit bir karar verme işlemini gerçekleştirmektedir (Safavian ve ark., 1991). Ağaç modellerinde bağımlı değişken kategorik yapıda olduğu durumlarda sınıflama ağacı, bağımlı değişken sürekli bir değişken olduğu durumlarda ise regresyon ağacı modeli kullanılmaktadır (Sümbüloğlu ve Akdağ, 2007; Nefeslioğlu ve ark., 2010).

Sınıflama ve Regresyon ağaçları bağımsız değişkene ait hiçbir ön koşul öne sürmeden kesikli ya da sürekli bağımlı değişkenin sınıf üyeliğini tahmin etmeye yarayan ters ağaç şeklindeki modellerdir. Kategorik ya da sürekli, bir ya da birden fazla bağımsız değişkenin kombinasyonları kullanılarak, tekrarlamalı ikili homojen bö-

lünmelerle, bağımlı değişkendeki değişimi ortaya çıkarmaya ve bağımlı değişkenin değerlerini tahmin etmeye yarayan ve görsel olarak ters ağaç şeklindeki modellere ağaç modelleri denir. İstatistik verilerin görsel olarak sunulması, aralarındaki etkileşimin belirlenmesi için karar ağaç modelleri sıklıkla kullanılmaktadır. Ağaç modellerinin işleyiş yapısı, bağımsız değişkene ait temel basit sorulardan alınan cevapların yarattığı yolları (ağaç dalları) takip etmektedir. Bu yollar ise (ağaç dalları) bağımlı değişkeni hangi bağımsız değişken ya da değişkenlerin etkilediğini gösterir (Temel, 2004).

Sınıflama ve Regresyon Ağacı yöntemi bazı araştırmacılar tarafından süt sığırcılığında (Doğan, 2003; Bakır ve ark., 2009; Bakır ve ark., 2010; Topal ve ark., 2010; Oruçoğlu, 2011; Çak ve ark., 2013; Eydurun ve ark., 2013b; Yılmaz ve ark., 2013), koyunculukta (Eydurun ve ark., 2008; Mohammad ve ark., 2012), tavukçulukta (Mendeş ve ark., 2009; Küçüköğlü, 2010) ve keçicilikte (Eydurun ve ark., 2013a) kullanılmıştır. Bunun yanında Sınıflama ve Regresyon Ağacı yönteminin sosyoloji (Oğuzlar, 2004; Kayri ve Boysan, 2008), tıp (Temel, 2004; Kıran, 2010), iktisat (Emel ve Taşkın, 2005; Avcı ve Altay, 2014) alanlarında da kullanıldığı görülmektedir.

Bu çalışmada, besi sonu canlı ağırlığını (BSCA) etkileyen çeşitli vücut ölçülerinin ((cidago yüksekliği (CY), sırt yüksekliği (SIY), besi süresi (BS), sağrı yüksekliği (SAY), kuyruk sokumu yüksekliği (KSY), vücut uzunluğu (VU), arka sağrı genişliği (ASG), göğüs derinliği (GD), ve göğüs çevresi (GÇ)) sınıflama ve regresyon ağacı yöntemiyle incelenmesi amaçlanmıştır.

2. Materyal ve Yöntem

2.1. Materyal

Araştırmanın hayvan materyalini Afyonkarahisar İli Bolvadin İlçesinde bulunan Avşar Tarım ve Hayvancılık İşletmesinde yetiştirilen 38 baş Siyah Alaca, 27 baş Simental, 23 baş Esmer ve 15 baş Melez erkek sığırlar oluşturmuştur. Bu hayvanlardan çeşitli vücut ölçüleri (cidago yüksekliği (CY), sırt yüksekliği (SIY), besi süresi (BS), sağrı yüksekliği (SAY), kuyruk sokumu yüksekliği (KSY), vücut uzunluğu (VU), arka sağrı genişliği (ASG), göğüs derinliği (GD), ve göğüs çevresi

(GÇ) ve canlı ağırlık değerleri (besi sonu canlı ağırlık (BSCA)) elde edilmiştir.

Ele alınan parametreleri ölçmek için 0.5 cm hassasiyetinde ölçü bastonu ve 0.5 cm hassasiyetinde ölçü şeridi, tartmak için ise 1 kg hassasiyetinde canlı hayvan kantarı (baskül) kullanılmıştır. Besi süreleri ile ilgili bilgiler işletmede bulunan kayıt defterinden alınmıştır.

2.2. Yöntem

Sınıflama ve regresyon ağacı, bütün bağımsız değişkenleri kullanıp verileri alt gruplara ayırarak oluşturulan bir ağaçtır. Sınıflama ve regresyon ağaçlarının en başında, herhangi bir parçalanma içermeyen ve bağımlı değişkenin yer aldığı kök düğümü bulunur. İlk olarak bu kök düğümü iki parçaya ayrılır. Bu iki parçaya, ebeveyn dalı adı verilir. Regresyon ağacının oluşturulmasında temel ilke, cevap değişkeninde maksimum homojenliği ya da saflığı sağlayacak şekilde tekrarlamalı olarak iki yavru düğüme parçalanmasıdır. Parçalanma sonucu oluşan düğümler, alt küme olarak da adlandırılır (Keskin ve ark., 2007). Ağacın oluşturulmasındaki temel amaç, cevap değişkenlerinde tekrarlamalı olarak oluşturulan herhangi bir yavru düğümden homojenlik mümkün olduğunca sağlanmışsa, bu düğümlerde artık parçalanması süreci sone erer ve bu düğüm terminal ya da uç düğüm olarak adlandırılır (Oruçoğlu, 2011). Bu süreçte; program, modele alınan bütün açıklayıcı değişkenleri test ederek, sonuçta yeni oluşacak olan düğümden, en yüksek homojenliği sağlayacak şekilde açıklayıcı değişkenin kesim değerini (eğer açıklayıcı değişken kategorik ise kategoriyi) belirler (Keskin ve ark., 2007).

İlk defa Morgan ve Sonquist (1963) tarafından ortaya atılan (Orekici, 2004), Breiman ve ark., (1984) tarafından yazılan "Classification and Regression Trees" isimli eser ile gelişme sağlayan sınıflandırma ve regresyon ağaçları, veri madenciliğinin sınıflandırma ile ilgili konuları arasında yer almaktadır. Sınıflandırma ağaçlarında kullanılacak birçok safsızlık ölçüsü (Gini, Twoing, Ordered Twoing ve Least Squared Deviation) bulunmaktadır (Kurt ve ark., 2008). Burada amaç hedef değişkeni açısından mümkün olduğunca homojen veri kümeleri üretmektir.

Regresyon ağaçlarında ise sınıflar yoktur. Bu nedenle regresyon ağacında Gini indeksi kullanılarak ayırma yapılamaz. Regresyon ağacındaki ayrımlar sonuçlanan iki düğüm için tahmin edilen toplam varyansın minimize olmasının gerekliliğini ifade eden "artıkların karelerini azaltma algoritması"na göre gerçekleştirilir (Breiman ve ark., 1984; Özkan, 2012).

Sınıflama ve regresyon ağaçlarını oluşturabilmek için ağaca önce büyüme daha sonra budama işlemi uygulanır (Küçüköğlu, 2010). Regresyon ağacı analizinin budama sürecinde yüksek hatalı sınıflandırmaya neden olan diğer bir ifade ile modelin tahmin edici değerine az katkıda bulunan düğümler ya da dallar atılmaktadır. Budama işlemi, cost-complexity parametresine göre yapmakta, bu parametre, değer ya da maliyet fonksiyonu,

hatalı sınıflandırma ve ağaç büyüklüğü (regresyon ağacındaki düğüm sayısı) değerleri dikkate alınarak hesaplanmaktadır (Keskin ve ark., 2007). Budama işlemi, ağaçta oluşmuş fakat sonucu etkilemeyen ve sınıflanmaya herhangi bir katkısı olmayan dalların ağaçtan alınmasıyla bir bakıma ağacın gereksiz ayrımlardan arındırılmasıdır. Budama uygulanmasının amacı önceki aşamalarda modele dahil edilmiş değişkenlerin ileriki aşamalarda tekrar modele dahil olmasını engellemektir. Budama süreci, en az katkı sağlayan düğümden başlayarak, önemli katkı sağlayan düğümler kalıncaya kadar devam ettirilir. Budanmış ağaç diğer yarısına gelene kadar tekrar budanır. Bu budama işlemi ağacın boyu artık değişmeye kadar devam eder (Küçüköğlu, 2010).

Regresyon ağacı yöntemi ile sınıflama ağacı için hatalı sınıflandırma oranı, regresyon ağacı içinde model tarafından açıklanamayan varyans yani hata varyansının hesaplanabileceğini, dolayısıyla, regresyon analizindeki R^2 'ye benzer ölçü ile modelin belirleme katsayısı hesaplanabilmektedir (Keskin ve ark., 2007).

Regresyon ağaçları, sınıflandırmaya dayalı en çok kullanılan model olmasındaki en büyük etken, kullanımı ve sonuçlarının anlaşılır olmasıdır. Temel olarak regresyon ağaçlarından sonuç elde etmenin 2 aşaması bulunmaktadır. Birincisi ağacın kurulmasıdır. Regresyon ağacının kurulmasında, daha hızlı ve güvenli bir sonuç alınması için verileri en büyük parçaya bölecek sorular kök düğümden itibaren sorulmaya başlanmalıdır. Regresyon ağacı yapısı oluşturulduktan sonra, ikinci aşamaya geçilip elde edilen veri, ağacın uygun dalına yerleştirilir.

Bir ağaç modelinde Y bağımlı, X_1 ve X_2 ise bağımsız değişken olarak ele alınırsa, X_1 ve X_2 ; $[-1; +1]$ tanım aralığında değişen düzgün (uniform) olasılık dağılımından tesadüfi olarak seçilen n_1 ve n_2 büyüklükteki örneklerin içerdiği değişken değerlerine çarpma kuralı uygulanarak aşağıdaki sonuçlar elde edilir. Bu sonuçlara göre;

$$X_1 \cdot X_2 \geq 0 \text{ ise } Y' \text{ nin değeri pozitif,}$$

$$X_1 \cdot X_2 \leq 0 \text{ ise } Y' \text{ nin değeri negatiftir.}$$

Bu şekilde bağımlı değişkenin pozitif ve negatif olmak üzere iki seviyesi oluşur (Temel, 2004).

Ağaç modellerinde karar verme noktalarına düğüm denmektedir. Ağaç modelinde başlangıç düğümü, gözlem değerlerinin tümünü ihtiva eden ve en karmaşık düğüm olan Düğüm I'dir. Bu düğüme kök düğümü ya da aile düğümü de denir. Kök düğümü iki alt düğüme (çocuk düğümüne) bölünür. Çocuk düğümlerinde henüz karar verme gerçekleşmemiştir. Aile düğümünden her çocuk düğümüne bölünme gerçekleştiği için çocuk düğümü aile düğümüne göre daha homojendir. Daha sonra çocuk düğümleri ayırma kriterleri dikkate alınarak karar noktalarına yani terminal düğümlere ulaşılır. Terminal düğümlerde ele alınan özelliklerin sınıf üyelikleri tanımlanır. Terminal düğümler ağaçtaki en homojen düğümler olduğu için, daha sonra bölünme gerçekleşmez (Temel, 2004).

Ağaç modellerinde, başlangıç düğümünden başlayarak ikili tekrarlı ayırmalarla daha homojen alt gruplara ulaşıp karar noktalarında bağımlı değişkenin durumu tanımlanmaktadır. Bu şekilde regresyon ağaçlarındaki düğüm noktalarında yer alan gözlemler sahip oldukları bağımsız değişkenin değerlerine göre iki çocuk düğümünden uygun olana atanırlar. Düğüm I'deki gözlemler eğer sıfırdan küçük veya sıfıra eşit iseler Düğüm II'ye, sıfırdan büyük iseler Düğüm III'e atanırlar. Buna benzer şekilde Düğüm II'deki gözlemler bağımsız değişkenin aldığı değere göre Düğüm IV ve Düğüm V'e, Düğüm III'deki değerler ise Düğüm VI ve Düğüm VII'ye atanırlar (Temel, 2004).

Bu çalışmada, CHAID algoritması için ebeveyn ve çocuk düğümlerinde minimum tosun sayıları 8:4 olarak ayarlanmıştır. CHAID algoritmasının etkinliğini belirlemek amacıyla model değerlendirme kriterleri olarak belirleme katsayısı, düzeltilmiş belirleme katsayısı ve gerçek ve tahmin edilen BSCA değerleri arasındaki Pearson korelasyon katsayısı kullanılmıştır (Grzesiak ve Zaboriski, 2012).

Model değerlendirme kriterlerinin mümkün olduğunca yüksek olması, kullanılan CHAID algoritması ile özelleştirilen modelin daha iyi olması anlamına gelmektedir.

3. Araştırma Sonuçları ve Tartışma

Çalışmada kullanılan 38 baş Siyah Alaca, 23 baş Esmer, 27 baş Simental ve 15 baş Melez erkek sığırlara ait çeşitli vücut ölçüleri (ciddago yüksekliği (CY), sırt yüksekliği (SIY), besi süresi (BS), sağrı yüksekliği (SAY), kuyruk sokumu yüksekliği (KSY), vücut uzunluğu (VU), arka sağrı genişliği (ASG), göğüs derinliği (GD), ve göğüs çevresi (GÇ)) ve canlı ağırlık değerlerine (besi sonu canlı ağırlık (BSCA)) ait tanıtıcı istatistikler Tablo 1'de verilmiştir.

Tablo 1'den de görülebileceği gibi her bir ırk için besi sonu canlı ağırlık değerleri sırasıyla 454.1, 455.9, 458.7 ve 429.2 kg olarak gerçekleşmiştir.

Yapılan varyans analizi neticesinde Siyah Alaca, Esmer, Simental ve melez erkek sığırlara ait BSCA, BS, VU, ASG, GD ve GÇ değerleri arasındaki farklar istatistik olarak önemsiz bulunmuştur ($P > 0.05$). Irklar arasında CY bakımından % 5'e, SIY, SAY ve KSY için ise % 1'e göre önemli farklılıklar tespit edilmiştir.

Ekonomik öneme sahip BSCA özelliğini etkileyen değişkenleri belirlemek amacıyla CHAID algoritması ile oluşturulan regresyon ağacı diyagramı Şekil 1'de sunulmuştur. Regresyon ağacıma ait hesaplanan model kalite değerleri sırasıyla, % 87.82 R^2 , % 87.32 Düz. R^2 ve gerçek ve tahmin edilen BSCA değerleri arasındaki korelasyon 0.937 olarak hesaplanmıştır. Bu değerler, BSCA özelliğinde meydana gelen toplam varyasyonun yaklaşık % 90'ının GÇ, VU, BS ve KSY bağımsız değişkenleri tarafından açıklandığını ifade etmektedir. Oluşturulan regresyon ağacının en tepesinde bulunan ve kök düğüm olarak adlandırılan düğüm, analize dahil edilen

bütün hayvanların bulunduğu grubu temsil etmektedir. Regresyon ağacı diyagramı incelendiğinde, BSCA üzerinde birinci derecede etkili bağımsız değişkenin GÇ (Düz. $P=0.000$), ikinci derecede etkili bağımsız değişkenlerin VU (Düz. $P=0.026$) ve BS (Düz. $P=0.000$), üçüncü derecede etkili bağımsız değişkenin KSY (Düz. $P=0.004$) olduğu belirlenmiştir.

Üzerinde çalışma yapılan tüm sığırlar (Düğüm 0), GÇ değişkeni bakımından 5 alt gruba (Düğüm 1, Düğüm 2, Düğüm 3, Düğüm 4 ve Düğüm 5) ayrılmıştır. Düğüm 1; göğüs çevresi $GÇ \leq 164$ cm olan tosunların oluşturduğu alt grubu, Düğüm 2; göğüs çevresi $164 < GÇ \leq 175$ cm arasında olan tosunların oluşturduğu alt grubu, Düğüm 3; göğüs çevresi $175 < GÇ \leq 183$ cm arasında olan tosunların oluşturduğu alt grubu, Düğüm 4; göğüs çevresi $183 < GÇ \leq 190$ cm arasında olan tosunların oluşturduğu alt grubu ve Düğüm 5; göğüs çevresi $GÇ \geq 190$ cm olan tosunların oluşturduğu alt grubu temsil etmektedir (Şekil 1). Düğüm 1 (göğüs çevresi $GÇ \leq 164$ cm olan tosunların) BSCA ortalaması 317.556 ($S=24.099$) kg, Düğüm 2 (göğüs çevresi $164 < GÇ \leq 175$ cm arasında olan tosunların) BSCA ortalaması 395.682 ($S=31.077$) kg, Düğüm 3 (göğüs çevresi $175 < GÇ \leq 183$ cm arasında olan tosunların) BSCA ortalaması 443.966 ($S=33.014$) kg, Düğüm 4 (göğüs çevresi $183 < GÇ \leq 190$ cm arasında olan tosunların) BSCA ortalaması 487.435 ($S=37.113$) ve Düğüm 5 (göğüs çevresi $GÇ \geq 190$ cm olan tosunların) BSCA ortalaması BSCA 545.850 ($S=28.087$) olarak bulunmuştur (Şekil 1).

Oluşturulan alt gruplardan yeterince homojen yapıya sahip olan Düğüm 1, Düğüm 2 ve Düğüm 5'e terminal düğüm adı verilir. Düğüm 1'den Düğüm 5'e doğru gidildikçe yani GÇ arttıkça BSCA ortalama değeri 317.556 kg'dan 545.850 kg'a artmıştır.

Düğüm 3 (göğüs çevresi $175 < GÇ \leq 183$ cm arasında olan tosunların) BSCA, VU tarafından etkilenmiştir (Düz $P=0.026$). Düğüm 3 (göğüs çevresi $175 < GÇ \leq 183$ cm arasında olan tosunların oluşturduğu alt grup) VU bakımından iki yeni alt gruba (Düğüm 6 ve Düğüm 7) ayrılmıştır. Düğüm 6; göğüs çevresi $175 < GÇ \leq 183$ cm ve vücut uzunluğu $VU \leq 135$ cm olan tosunların oluşturduğu alt grubu, Düğüm 7; göğüs çevresi $175 < GÇ \leq 183$ cm ve vücut uzunluğu $VU > 135$ cm olan tosunların oluşturduğu alt grubu ifade etmektedir. Düğüm 6 ve Düğüm 7 için BSCA ortalamaları sırasıyla, 406.000 ($S=17.819$) kg ve 451.875 ($S=29.883$) kg olarak tahmin edilmiştir. Düğüm 7 (göğüs çevresi $175 < GÇ \leq 183$ cm ve vücut uzunluğu $VU > 135$ cm olan tosunlar) BSCA sadece KSY tarafından etkilendiği saptanmıştır (Düz. $P=0.004$).

Düğüm 7 (göğüs çevresi $175 < GÇ \leq 183$ cm, vücut uzunluğu $VU > 135$ cm olan tosunların oluşturduğu alt grup) KSY bakımından iki yeni alt gruba (Düğüm 10 ve Düğüm 11) ayrılmıştır. Düğüm 10; göğüs çevresi $175 < GÇ \leq 183$ cm, vücut uzunluğu $VU > 135$ cm ve kuyruk sokumu yüksekliği $KSY \leq 132$ cm olan tosunların oluşturduğu alt grubu, Düğüm 11; göğüs çevresi $175 < GÇ$

≤ 183 cm, vücut uzunluğu $VU > 135$ cm ve kuyruk sokumu yüksekliği $KSY > 132$ cm olan tosunların oluşturduğu alt grubu ifade etmektedir.

Tablo 1

Çeşitli vücut ölçüleri ve ırklara göre tanıtıcı istatistikler

| Değişkenler | İrk | N | $\bar{X} \pm S_{\bar{X}}$ | Minimum | Maximum | CV |
|-------------|-----|----|----------------------------|---------|---------|-------|
| BSCA (Kg) | 1 | 38 | 454.1±12.4 | 276 | 571 | 16.80 |
| | 2 | 23 | 455.9±17.6 | 285 | 588 | 18.47 |
| | 3 | 27 | 458.7±13.4 | 313 | 587 | 15.20 |
| | 4 | 15 | 429.2±12.5 | 335 | 521 | 11.29 |
| BS (Gün) | 1 | 38 | 4.763±0.122 | 4 | 6 | 15.77 |
| | 2 | 23 | 4.957±0.194 | 4 | 6 | 18.73 |
| | 3 | 27 | 4.815±0.142 | 4 | 6 | 15.28 |
| | 4 | 15 | 4.467±0.215 | 4 | 6 | 18.67 |
| CY (cm) | 1 | 38 | 129.11±1.11 ^a | 111 | 144 | 5.29 |
| | 2 | 23 | 125.80±1.21 ^{ab} | 110 | 134.5 | 4.61 |
| | 3 | 27 | 125.89±0.955 ^{ab} | 119 | 135 | 3.94 |
| | 4 | 15 | 123.63±1.10 ^b | 116 | 133 | 3.43 |
| SIY (cm) | 1 | 38 | 131.76±0.984 ^A | 114.5 | 146 | 4.60 |
| | 2 | 23 | 128.96±1.19 ^{AB} | 118 | 138 | 4.41 |
| | 3 | 27 | 128.11±1.03 ^{AB} | 119 | 140 | 4.18 |
| | 4 | 15 | 125.37±1.27 ^B | 117.5 | 134 | 3.91 |
| SAY (cm) | 1 | 38 | 134.84±1.06 ^A | 116 | 149 | 4.83 |
| | 2 | 23 | 131.52±1.15 ^{AB} | 118 | 143 | 4.19 |
| | 3 | 27 | 130.39±0.992 ^{AB} | 123 | 141 | 3.95 |
| | 4 | 15 | 128.67±1.16 ^B | 118 | 135 | 3.49 |
| KSY (cm) | 1 | 38 | 134.96±0.936 ^A | 117 | 147 | 4.27 |
| | 2 | 23 | 131.37±0.934 ^{AB} | 120 | 140 | 3.41 |
| | 3 | 27 | 130.09±0.774 ^B | 124 | 141 | 3.09 |
| | 4 | 15 | 129±0.975 ^B | 121 | 135 | 2.93 |
| VU (cm) | 1 | 38 | 143.24±1.34 | 124 | 160 | 5.79 |
| | 2 | 23 | 141.39±2.01 | 122 | 163 | 6.83 |
| | 3 | 27 | 141.70±1.31 | 128 | 155 | 4.80 |
| | 4 | 15 | 141.40±1.52 | 129 | 153 | 4.16 |
| ASG (cm) | 1 | 38 | 48.105±0.615 | 40 | 59 | 7.88 |
| | 2 | 23 | 49.04±1 | 38 | 58 | 9.81 |
| | 3 | 27 | 49.519±0.746 | 41 | 55 | 7.83 |
| | 4 | 15 | 48.8±0.947 | 44 | 57 | 7.52 |
| GD (cm) | 1 | 38 | 62.842±0.581 | 53 | 69 | 5.70 |
| | 2 | 23 | 61.96±1.07 | 50 | 71 | 8.31 |
| | 3 | 27 | 61.815±0.710 | 55 | 69 | 5.97 |
| | 4 | 15 | 61.933±0.808 | 56 | 67 | 5.05 |
| GÇ (cm) | 1 | 38 | 181.12±1.79 | 155 | 197 | 6.11 |
| | 2 | 23 | 180.22±2.85 | 154 | 203 | 7.58 |
| | 3 | 27 | 181.44±1.66 | 162 | 197 | 4.75 |
| | 4 | 15 | 179.27±2.02 | 164 | 191 | 4.37 |

^{a, b}: $P < 0.05$; ^{A, B}: $P < 0.01$

Düğüm 4 (göğüs çevresi $183 < GÇ \leq 190$ cm aralığında olan tosunların) BSCA özelliği üzerinde analize dahil edilen bağımsız değişkenlerden sadece BS etkili bulunmuştur (Düz. $P=0.000$). Düğüm 4; BS bakımından iki yeni alt gruba (Düğüm 8 ve Düğüm 9) ayrılmıştır. Düğüm 8; göğüs çevresi $183 < GÇ \leq 190$ cm ve besi süresi $BS \leq 4$ olan tosunların oluşturduğu alt grubu, Düğüm 9; göğüs çevresi $183 < GÇ \leq 190$ cm ve besi süresi $BS > 4$ olan tosunların oluşturduğu alt grubu ifade etmektedir. Düğüm 8 ve Düğüm 9 için BSCA ortalamaları

sırasıyla, 446.000 ($S=32.573$) kg ve 505.562 ($S=21.071$) kg olarak hesaplanmıştır. Göğüs çevresi $183 < GÇ \leq 190$ cm tosunların 4 aydan daha fazla sürede beside tutulması BSCA ortalaması olarak arttırdığı söylenebilir.

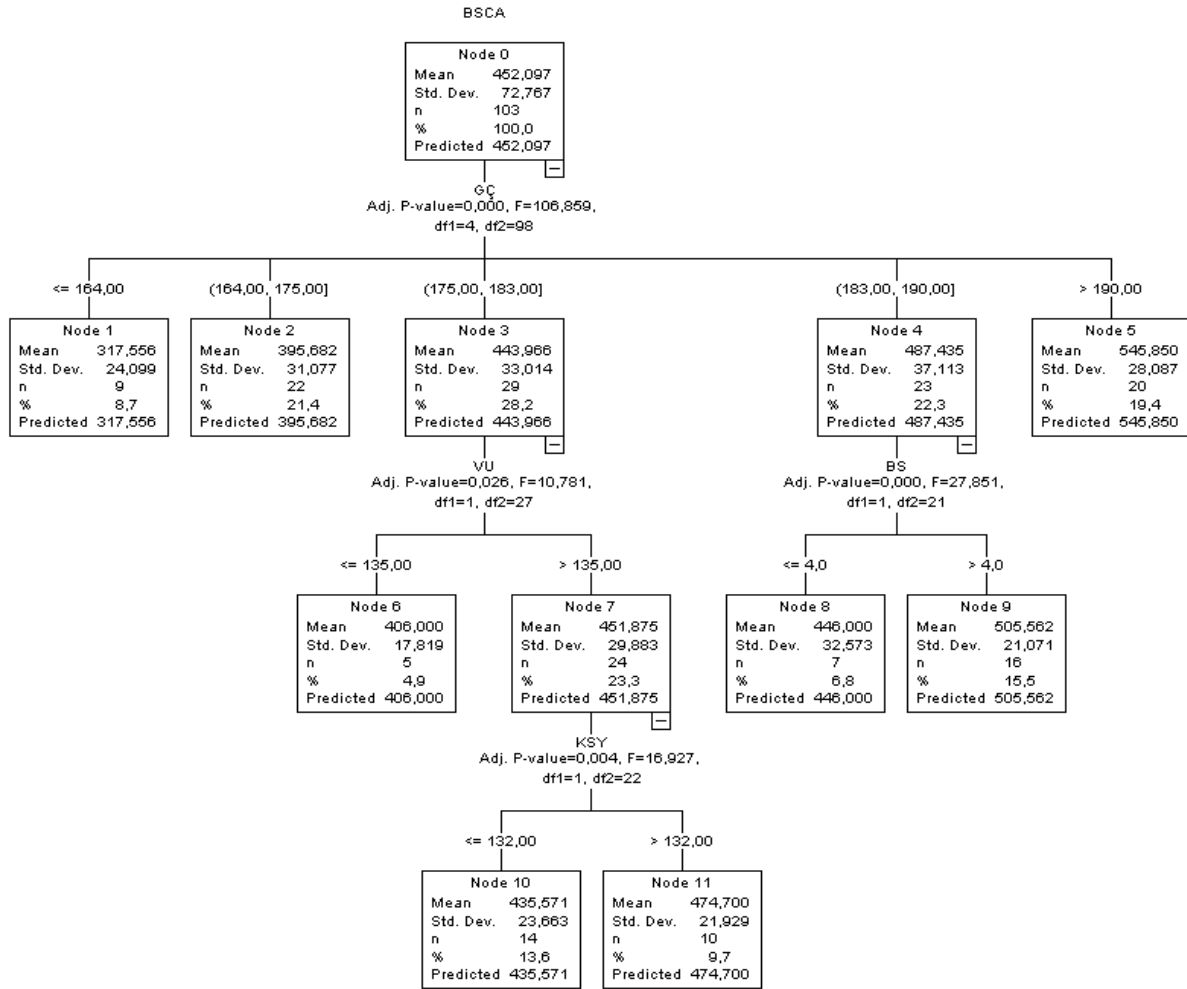
4. Sonuçlar ve Öneriler

Klasik sınıflama modellerine alternatif olarak ortaya çıkan regresyon ağacı, modelde yer alan değişkenler

üzerinde hiçbir varsayım gerektirmemesi nedeniyle, eldeki verilerin fazla sayıda ve kompleks bir yapıda olduğu durumlarda sınıflama analizlerinde kolaylıkla uygulanabilen güçlü bir tekniktir. Regresyon ağacı analiziyle oluşturulan diyagram incelendiğinde bağımlı değişkenleri hangi bağımsız değişkenlerin etkilediği kolaylıkla görülebilmektedir.

En ağır BSCA (545.850 kg); göğüs çevresi GÇ > 190 cm olan tosunların oluşturduğu alt gruptan elde edilmiştir. İkinci olarak en ağır BSCA (505.562 kg); 183 < GÇ ≤ 190 cm ve BS > 4 ay olan tosunların oluşturduğu alt gruptan elde edilmiştir.

Sonuç olarak, farklı vücut ölçülerinden besi sonucu canlı ağırlığı tahmin etmek amacıyla CHAID algoritmasının kullanılması tavsiye edilebilir.



Şekil 1

Regresyon ağacı diyagramı

5. Teşekkür

Bu çalışma Rifat Akşahan'ın Yüksek Lisans tez çalışmasından üretilmiştir.

6. Kaynaklar

Avcı MA, Altay NO (2014). Finansal Krizlerin Öngörüşünde Regresyon Ağaçları Modeli:

Gelişmekte Olan Ülkelere Yönelik Bir Analizi. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 12: 191-212.

Bakır G, Keskin S, Mirtagioğlu 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, Mirtagioğlu 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.
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984). Classification and Regression Trees, New York, Chapman and Hall.
- Çak B, Keskin S, Yılmaz O (2013). Regression Tree Analysis for Determining of Affecting Factors to Lactation Milk Yield in Brown Swiss Cattle, *Asian Journal of Animal and Veterinary Advances*, 8 (4): 677-682.
- Doğan İ (2003). Holştayn Irkı İneklerde Süt Verimine Etki Eden Faktörlerin CHAID Analizi İle İncelenmesi. *Ankara Üniversitesi, Veteriner Fakültesi Dergi*, 50: 65-70.
- Efe E, Bek Y, Şahin M (2000). SPSS'te Çözümleri ile İstatistik Yöntemler II. Kahramanmaraş: Kahramanmaraş Sütçü İmam Üniversitesi Rektörlüğü, Yayın No:10.
- Emel GG, Taşkın Ç (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6 (2): 221-239.
- 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, Kaygisiz A, Aktaş ZM (2013a). 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 of Animal & Plant Sciences*, 23(4): 956-963.
- Eyduran E, Yılmaz I, Kaygisiz A, Tariq MM., (2013b). Estimation of 305-D Milk Yield Using Regression Tree Method in Brown Swiss Cattle. *The Journal of Animal & Plant Sciences*, 23(3): 731-735.
- Grzesiak W, Zaborski D (2012). Example of the Use of Data Mining Methods in Animal Breeding. ISBN: 978-953-51-0720-0.
- Kayri M, Boysan M (2008). Bilişsel Yatkınlık İle Depresyon Düzeyleri İlişkisinin Sınıflandırma ve Regresyon Ağacı ile İncelenmesi. *Hacettepe Üniversitesi Eğitim Bilimleri Dergi*, 34: 168-177.
- Keskin S, Kor A, Karaca S (2007). Use of Factor Analysis Scores in Multiple Linear Regression Model for Determining Relationships Between Milk Yield and Some Udder Traits in Goats. *Journal of Applied Animal Research*, 31: 185-188.
- Kıran ZB (2010). Lojistik Regresyon ve Cart Analizi Teknikleriyle Sosyal Güvenlik Kurumu İlaç Provizyon Sistemi Verileri Üzerinde Bir Uygulama. Yüksek Lisans Tezi, *Gazi Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Bölümü*, Ankara.
- Kurt I, Ture M, Kurum AT (2008). Comparing Performances of Logistic Regression, Classification and Regression Tree, and Neural Networks for Predicting Coronary Artery Disease. *Expert Systems with Applications*, 34: 366-374.
- Küçüköğlü O (2010). Veri madenciliği yöntemlerinin hayvancılıkta kullanımı. Yüksek Lisans Tezi, *Çanakkale Onsekiz Mart Üniversitesi, Fen Bilimleri Enstitüsü*, Çanakkale.
- Mendeş M, Akkartal E (2009). Regression Tree Model in Predicting Slaughter Weight of Broiler, *Italian Journal of Animal Science*, 8: 615-624.
- Mohammad MT, Rafeeq M, Bajwa, MA, Abbas F, Waheed A, Bukhari FA, 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 Sciences*, 22 (1): 20-24.
- Morgan JN, Sonquist JA (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of The American Statistical Association* 58; 415-434.
- Nefeslioglu HA, Sezer E, Gokceoglu C, Bozkir AS, Duman TY (2010). Assessment of Landslide Susceptibility by Decision Trees in the Metropolitan Area of Istanbul, Turkey. *Mathematical Problems in Engineering*, doi:10.1155/2010/901095.
- Oğuzlar A (2004). CART Analizi ile Hane halkı İşgücü Anketi Sonuçlarının Özetlenmesi, *İktisadi ve İdari Bilimler Dergisi*, 18 (3-4): 79-90.
- 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.
- Özkan K (2012). Sınıflandırma ve Regresyon Ağacı Tekniği (SRAT) ile Ekolojik Verinin Modellenmesi, *SDÜ Orman Fak. Dergisi*, 13: 1-4.
- Safavian SR, Landgrebe D (1991). A Survey of Decision Tree Classifier Methodology, *IEEE Transactions on Systems Man and Cybernetics*, 21: 660-674.
- Sümbüloğlu K, Akdağ B (2007). Regresyon Yöntemleri ve Korelasyon Analizi, *Hatiboğlu Yayınevi*. Ankara.
- Temel ÖG (2004). Sınıflama ve Regresyon Ağaçları, Yüksek Lisans Tezi, *Mersin Üniversitesi, Sağlık Bilimleri Enstitüsü*, Mersin.
- Topal M, Aksakal V, Bayram B, Yaganoglu M (2010). An Analysis of the Factor Affecting Birth Weight and Actual Milk Yield in Swedish Red Cattle Using Regression Tree analysis. *The Journal of Animal & Plant Sciences*, 20: 63-69.
- Yılmaz I, Eyduran E, Kaygisiz A (2013). Determination of Non-Genetic Factors Influencing Birth Weight Using Regression Tree Method in Brown-Swiss Cattle Canadian. *Journal of Applied Science*, 1 (3): 382-387.