



**HAYVANCILIK ALANINDA ÇOK DEĞİŞKENLİ  
UYARLANABİLİR REGRESYONUN  
KULLANIMI**

**Orhan AKKUŞ**

**Yüksek Lisans Tezi**

**ZOOTEKNİ ANABİLİM DALI**

**Danışman: Doç. Dr. Ecevit EYDURAN**

**2017**

**T.C**

**İĞDIR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ**

**HAYVANCILIK ALANINDA ÇOK DEĞİŞKENLİ UYARLANABİLİR  
REGRESYONUN KULLANIMI**

**Orhan AKKUŞ**

**ZOOTEKNİ ANABİLİM DALI**

**İĞDIR**

**2017**

**Her hakkı saklıdır**

Doç. Dr. Ecevit EYDURAN danışmanlığında Orhan AKKUŞ tarafından hazırlanan bu çalışma ..... tarihinde aşağıdaki jüri üyeleri tarafından Zootekni Anabilim Dalı'nda Yüksek lisans tezi olarak kabul edilmiştir.

**Başkan** : Doç. Dr. Ecevit EYDURAN

İmza:

**Üye** : Yrd. Doç. Dr. Mehmet Kazım KARA

İmza:

**Üye** : Yrd.Doç.Dr.Şenol ÇELİK

İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ..... / ..... /2017 tarih ve 2017/..... sayılı kararı ile onaylanmıştır.

(İmza)

.....

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.

Orhan AKKUŞ



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

# HAYVANCILIK ALANINDA ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYONUN KULLANIMI

AKKUŞ, Orhan

Yüksek Lisans Tezi, Zootekni Anabilim Dalı

Tez Danışmanı: Doç. Dr. Ecevit EYDURAN

Eylül 2017, 30 Sayfa

Bu çalışmada, hayvancılık ile ilgili bir veri setinde MARS algoritmasının nasıl kullanılacağını ve elde edilen sonuçların nasıl yorumlanacağı amaçlanmıştır. Canlı ağırlığı tahmin etmek amacıyla Pakistan'da yetiştirilen dört koyun ırkının (Balochi (48), Harnai (48), Beverigh (47) ve Rakhashni (48) verileri kullanılmıştır. MARS veri madenciliği algoritmasının tanıtılması ve canlı ağırlık için faydalı tahmin denkleminin geliştirilmesi amacıyla, cidago yüksekliği, göğüs çevresi ve vücut uzunluğu gibi sürekli değişkenler bağımsız değişken olarak kullanılmıştır. R yazılımının "earth" paketi ile vücut ağırlığının tahmini için iki MARS uygulaması (26 ve 38 terim), interaksiyon derecesi 2 ve 10 çapraz geçerlilik temel alınarak yapılmıştır. MARS uygulamalarının tahmin doğruluğunu değerlendirmek için uyum kriterleri (GCV, RSS,  $GR^2$ ,  $R^2$ , düzeltilmiş  $R^2$  ve çapraz geçerlilik  $R^2$ 'deki gerçek ve öngörülen değerler arasındaki korelasyon katsayısı) tahmin edilmiştir. Uyum kriterleri her iki uygulama için sırasıyla (0.980 ve 0.990), GCV (0.610 ve 0.293), RSS (116 ve 56),  $GR^2$  (0.960 ve 0.980),  $R^2$  (0.960 ve 0.980), düzeltilmiş  $R^2$  (0.954 ve 0.975) ve çapraz geçerlilik  $R^2$  (0.922 ve 0.88) olarak tahmin edilmiştir. R yazılımında en küçük GCV'yi üreten etkileşim dereceleri ve uygun terim sayısının atanması, MARS algoritması için dikkate alınması gereken en önemli noktalar. Elde edilen sonuçlar, MARS uygulamalarının, ırk standartlarını ortaya koymak ve burada incelenen canlı ağırlık özelliğini olumlu etkileyen morfolojik özellikleri belirlemek için önemli bir seçenek olabileceğini göstermiştir.

**Anahtar Kelimeler:** Veri Madenciliği, MARS, GCV, Doğrusal olmayan Regresyon Tekniği

## ABSTRACT

### USE OF MULTIVARIATE ADAPTIVE REGRESSION SPLINES FOR ANIMAL HUSBANDRY

AKKUŞ, Orhan

Master Thesis, Department of Animal Science

Thesis Advisor: Assoc. Prof. Dr. Ecevit EYDURAN

September, 2017, 30 pages

The aim of this study was to illustrate how to use MARS algorithm at a data set from Animal Science and to interpret the achieved results. Data of four sheep breeds (Balochi (48), Harnai (48), Beverigh (47) and Rakhshni (48)) reared in Pakistan were used in the prediction of body weight. Withers height, chest girth and body length as continuous predictors were used with the aim of introducing MARS data mining algorithm and developing the useful prediction equation for body weight. Breed was employed as a nominal predictor. With the package "earth" of R software, two MARS applications (26 and 38 term numbers) for the prediction of body weight were made by taking a basis for interaction degrees of 2, and v-ten-fold cross validation. For evaluating predictive accuracy of the MARS applications, goodness of fit criteria i.e.  $r$  (Pearson correlation coefficient between the actual and predicted values in body weight, GCV, RSS,  $GR^2$ ,  $R^2$ , adjusted  $R^2$  and cross validation  $R^2$  were estimated. Goodness of fit criteria i.e.  $r$  (0.980 and 0.990), GCV(0.610 and 0.293), RSS(116 and 56),  $GR^2$  (0.960 and 0.980),  $R^2$  (0.960 and 0.980), adjusted  $R^2$  (0.954 and 0.975) and cross validation  $R^2$  (0.922 and 0.88) were estimated for both applications, respectively. Assignment of the suitable term number and interaction degrees producing the smallest GCV in R software is the most important points that are necessary to be taken into consideration for MARS algorithm. The achieved outcomes illustrated that MARS applications could be a notable option for proving breed standards and determining body measurements positively affecting the body weight of the sheep studied here.

**Key words:** Data Mining, MARS, GCV, Non-linear regression technique

## ÖNSÖZ ve TEŞEKKÜR

Diğer sınıflama yöntemleri ile karşılaştırıldığında değişkenlerin dağılımına ilişkin varsayımlara gerek duymayan veri madenciliği algoritmalarının çok avantajlı olduğu görülmektedir. Veri madenciliği analizinde kullanılacak değişkenler; sınıflayıcı (nominal), sıralı (ordinal) ve sürekli (continuous) yapıda olabilir. Diğer çok değişkenli modelleme yöntemleri araştırmacılara daha fazla data gerektirirken, veri madenciliği yöntemi araştırmacılara daha az girdi gerekmektedir. Ayrıca, genel doğrusal modeller kapsamında sürekli bir bağımlı değişkeni tahmin etmek amacıyla kullanılan ve görsel bir ağaç yapısı oluşturan CART, CHAID, Exhaustive CHAID veri madenciliği algoritmaları ile elde edilen sonuçların kolay yorumlanabilir olduğunu söylemek mümkündür. Buna karşın, benzer amaçlı kullanılabilen yapay sinir ağları (Artificial Neural Networks, ANNs) algoritmaları, gerek bağımlı ve bağımsız değişken arasındaki karmaşık ilişkileri belirlenmesinde gerekse bağımsız değişkenler arasındaki doğrusal olmayan ve interaksiyon etkilerinin tanımlanmasında iyi bir seçenek olmayabilir. Şimdiye kadar bahsedilen algoritmaların aksine çalışılan veri setleri üzerinde yüksek uyum iyiliği sağlayabilen MARS algoritması, araştırmacıların tahmin denklemi elde etmesine olanak sağlar. Gerek iktisat gerekse etki faktörü yüksek olan mühendislik alanlarında kabul gören MARS algoritması, genel doğrusal modeller ya da birden fazla sürekli bağımlı değişken içeren modeller kapsamında kullanılmaktadır.

Son yıllarda, MARS algoritması, hayvancılık alanında, binary logistik regresyon yapısındaki veriler için kullanıldığı tespit edilmiştir. Ancak, sürekli bağımlı bir değişken için tahmin denklemi geliştirmek amacıyla MARS algoritmasının kullanımı hayvancılık literatüründe yok denecek kadar sınırlıdır. Değişkenlerin dağılımı ile ilgili herhangi bir varsayım gerektirmemesi ve önemli bağımsız değişkenlerin arasındaki doğrusal olmayan ve yüksek dereceli interaksiyon etkilerinin irdelenebilmesi gibi avantajlar sayesinde hayvancılık alanından elde edilen verilerin MARS yöntemiyle değerlendirilmesi ve analiz sonuçlarının yorumlanması hayvansal verilerin değerlendirilmesine farklı bir bakış açısı sağlayacaktır. Bununla birlikte bu çalışmada hayvancılık alanında çok az kullanılan MARS modeli için sürekli yapıda bağımlı değişken (canlı ağırlık) kullanılmış, MARS modelinin üstünlüğü vurgulanmıştır. Diğer

çalışmalardan farklı olarak bu çalışmada mevcut verilerin analizi için R yazılımı kullanılmış ve analiz yapım aşamasındaki bazı püf noktalara değinilmiştir.

Bu yararlı tez konusunu bana öneren, bu tezi yazmamda bana yol gösteren, engin tecrübe ve tavsiyeleri ile beni destekleyen, maddi ve manevi daha doğruya ulaşmama yardımcı olan saygıdeğer danışman hocam Doç. Dr. Ecevit EYDURAN'a, Zootekni bölümü öğretim üyelerinden Yrd. Doç. Dr. İsa YILMAZ'a ve Yrd. Doç. Dr. M. Kazım KARA'a hocalarıma sonsuz teşekkürlerimi sunarım.

Ayrıca bu tez çalışmasında verilerinin kullanılmasına izin veren Prof. Dr. Mohammad Masood TARIQ ve Dr. Abdul WAHEED' e çok teşekkür ederim.

Orhan AKKUŞ

Haziran – 2017



## İÇİNDEKİLER

	<b>Sayfa No</b>
ÖZET .....	i
ABSTRACT.....	ii
ÖNSÖZ ve TEŞEKKÜR.....	iii
İÇİNDEKİLER .....	V
SİMGELER ve KISALTMALAR DİZİNİ .....	VI
ŞEKİLLER DİZİNİ .....	VII
ÇİZELGELER DİZİNİ .....	VIII
1.GİRİŞ .....	1
2. KAYNAK ÖZETLERİ .....	4
3. MATERYAL ve METOT .....	9
3.1. Materyal .....	9
3.2. Metot .....	9
4. BULGULAR ve TARTIŞMA.....	15
5. SONUÇ ve ÖNERİLER .....	26
6.KAYNAKLAR .....	28
ÖZGEÇMİŞ .....	31

## SİMGELER ve KISALTMALAR DİZİNİ

### Simgeler

<b>%</b> .....	Yüzde
<b>cm</b> .....	Santimetre
<b>g</b> .....	Gram
<b>Kg</b> .....	Kilogram
<b>Lt</b> .....	Litre
<b>mm</b> .....	Milimetre

### Kısaltmalar

<b>CA</b> .....	Canlı ağırlık
<b>CY</b> .....	Cidago yüksekliği
<b>CHAID</b> .....	Otomatik Ki-Kare Etkileşim Belirleme Analizi
<b>CART</b> .....	Sınıflandırma ve regresyon ağacı
<b>VU</b> .....	Vücut uzunluğu
<b>GC</b> .....	Göğüs çevresi
<b>R<sup>2</sup></b> .....	Belirleme katsayısı
<b>R<sup>2</sup>-düz. (%)</b> .....	Düzeltilmiş belirleme katsayısı
<b>SSoran</b> .....	Standart sapma oranı

## ŞEKİLLER DİZİNİ

### Sayfa No

Şekil 4.1. Seçilen modele ilişkin belirlenme katsayıları ve hata terimlerine ait grafikler.....	17
Şekil 4.2. Bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiler. ....	18
Şekil 4.3. Seçilen modele ilişkin belirlenme katsayıları ve hata terimlerine ait grafikler.....	19
Şekil 4.4. Bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiler .....	20
Şekil 4.2.1. Seçilen modele ilişkin belirlenme katsayıları ve hata terimleri grafikleri ....	23
Şekil 4.2.2. Bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiler. ....	24

## ÇİZELGELER DİZİNİ

	<b>Sayfa No</b>
<b>Çizelge 4.1.</b> MARS tahmin denklemine ait katsayılar ve terimler .....	16
<b>Çizelge 4.2.</b> MARS tahmin denklemine ait katsayılar ve terimler .....	22
<b>Çizelge 4.3.</b> Çoklu regresyon analizine ait varyans analizi tablosu.....	25



## 1.GİRİŞ

Genel anlamda regresyon analizi bir bağımlı değişken ile bir ya da birden fazla bağımsız değişken arasındaki ilişkinin ortaya konulduğu analiz tekniğidir. Regresyon analizinde amaç; kolay ölçülebilen özellikler yani değişkenlerden yola çıkarak zor ölçülebilen bağımlı değişken için tahmin denklemi geliştirmektir. Regresyon analizinde temel amaç; tahmin hatasını azaltan bağımsız değişkenleri ortaya koymaktır. Diğer bir ifadeyle; Bağımlı değişkenin gerçek değeri ile (regresyon analizi ile tahmin edilen) tahmin değeri arasındaki korelasyonu arttırmak ya da farkı azaltmaktır. Çoklu doğrusal regresyon analizinde regresyon katsayılarına ilişkin parametrelerin tahmin edilmesinde kullanılan En Küçük Kareler Metodu (Least Squares Method=Ordinary Least Squares) kullanılmaktadır. Ancak çoklu doğrusal regresyon analizinde bazı temel varsayımların yerine gelmemesi durumunda, bu analizden elde edilen EKK parametre tahminleri güvenilir olmamaktadır. Örneğin, bağımsız değişkenler arasında yüksek korelasyonun olması sonucunda çoklu doğrusal regresyon analizinde çoklu bağlantılık (Multicollinearity) problemi oluşur. Bu durumda, çoklu doğrusal regresyon analizine alternatif yöntemlerin kullanılması mantıklı bir yaklaşım olacaktır. İncelenen verim üzerinde etkili olabilecek çevre faktörlerinin belirlenmesi ve verim özelliği üzerinde etkili diğer kantitatif özellikler arasındaki ilişkilerin ortaya konulması önem arz etmektedir (Koç, 2016). Bu ilişkilerin güvenilir olarak ortaya konulması, uygun istatistiksel metotlar kullanılmasına bağlıdır.

Çoklu doğrusal regresyon analizine alternatif olarak kullanılan istatistiksel metotlar; Ridge Regresyon Analizi (Ridge Regression Analysis), Stepwise Regresyon Analizi (Stepwise Regression Analysis), Faktör Analizi Skorları ile Çoklu Regresyon Analizi (Factor Analysis Scores in Multiple Linear Regression Analysis), Temel Bileşenler Analizi Skorları ile Çoklu Regresyon Analizi (Principal Component Analysis Scores in Multiple Linear Regression Analysis) ve Robust Regresyon Analizi (Robust Regression Analysis) olarak sıralanabilir. Bununla birlikte çoklu regresyon analizine alternatif olarak veri madenciliği algoritmaları da kullanılabilir. CART algoritması (Classification and Regression Tree), CHAID algoritması (Chi-Square Automatic Interaction Detector), Exhaustive CHAID algoritması (Exhaustive Chi-Square

Automatic Interaction Detector), Çok katmanlı perceptron (Multilayer Perceptron Neural Network Algorithm), ve radyan temel fonksiyonlu (Radial Basis Function Neural Network Algorithm) yapay sinir ađ algoritmaları ve ayrıca MARS algoritması bu kapsamda kullanılan veri madenciliđi algoritmalarıdır. Regresyon analizi yani sürekli bir deđişkenin tahmin edilmesi kapsamında yukarıda verilen ve yapay sinir ađları kapsamındaki veri madenciliđi algoritmaları sınıflayıcı (nominal), sıralı (ordinal) ve sürekli (continuous) bađımsız deđişkenler için rahatlıkla kullanılabilir. Bu nedenle, Nisbet et al. (2009), SPSS, STATISTICA ve SAS paket programlarını kullanarak CART, CHAID, Boosted Classifiers ve Regression, MARS, Regresyon ve Sınıflama için Random Forest, Makine Öğrenme (Bayesian, Destek Vektörleri, En yakın komşuluk, ve Bađımsız Bileşenler) gibi ileri veri madenciliđi algoritmalarını detaylı olarak çalışmışlardır. Çetin ve Mikail (2016), yapmış oldukları derleme çalışmasında; k ortalamaları yaklaşımı, k-en yakın komşu yaklaşımı, çok deđişkenli uyarlanır regresyon eğrileri (Multivariate Adaptive Splines, MARS), Bayes sınıflandırıcıları (Naive Bayesian Classifiers, NBC), yapay sinir ađları (Artificial Neural Networks, ANN), destek vektör makineleri (Support Vector Machines, SVM), karar ağaçları gibi çok yararlı veri madenciliđi algoritmaları konusunda teorik bilgiler vermişlerdir. Bu algoritmalar arasında MARS algoritması, Stanford'da fizikçi ve istatistikçi olan Jerome FRIEDMAN tarafından geliştirilmiştir. Bu algoritma; gerek bađımlı ve bađımsız deđişkenlerin dağılımlarına ilişkin herhangi bir varsayımının olmaması gerekse bađımlı deđişken ile bađımsız deđişkenler arasındaki (dođrusal olmayan ve yüksek dereceli interaksiyon etkileri şeklindeki) karmaşık ilişkileri ortaya koyarak tahmin denklemlerinin oluşturulmasına olanak sağlamaktadır. MARS algoritması birden fazla bađımlı ve bađımsız deđişken için eş zamanlı analiz yapma imkanına sahip olup yanıt yüzeyi metoduna (Response Surface Method) alternatif olarak kullanılabilir. Bununla birlikte MARS algoritması, yanıt yüzeyi metodu ile yapılan optimizasyon kapsamında azaltılmış faktöriyel deneme desenlerine (Reduced Factorial Designs) hem alternatif hem de bu desenler ile birlikte kullanılabilme imkanına sahiptir (Eyduran, 2016). MARS algoritması, tıp (Ankaralı et al. 2010; Zakeri et al. 2010), iktisat ve bankacılık (Tunay 2010; Tunay 2011; Acciani et al., 2011; Dođan ve Toprak 2012) ve Jeoloji mühendisliđi (Zhang et al. 2016) gibi farklı bilim alanlarında uygulama konusu olmuştur. Ancak literatürde regresyon analizi kapsamında

MARS modelinin hayvancılık alanında kullanımı yetersizdir (Eyduvan et al. 2017; Karadas et al. 2017). Bu eksikliđi gidermek amacıyla yrtlmş olan bu tez alıřmasında, hayvancılık alanından elde edilen bir veri setinin analizinde MARS algoritmasının R yazılım programı ile nasıl analiz edileceđi, MARS algoritmasının analizinde dikkat edilmesi gereken pf noktalardan ve elde edilen sonuların nasıl yorumlanacađından bahsedilecektir. Mevcut alıřmada uygulanan MARS analizi sonucunda elde edilen tahmin denklemi yardımıyla bađımlı deđiřken (canlı ađırlık) ile pozitif etkiye sahip bađımsız deđiřkenlerin birbiriyle oluřturduđu etkileřim hakkında bilgi sahibi olunacaktır.



## 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.

Grzesiak et al. (2003), kısmi laktasyon kayıtlarını kullanarak 305 günlük süt verimini tahmin etmeye çalışmışlardır. Araştırmacılar süt verimini tahmin etmek amacıyla yapay sinir ağları ve çoklu regresyon metotlarını kullanmışlardır. Modellerin karşılaştırılmasında, hata kareler ortalamasının karekökü, standart sapma oranı, tahmine ilişkin relative ortalama, hata ve düzeltilmiş belirleme katsayısı gibi model kalite ölçütleri kullanılmıştır. Çoklu regresyon modeline nazaran sinir ağları modelinin 305 günlük süt verimi gibi kantitatif özelliklerin tahmin edilmesinde alternatif bir metot olabileceği sonucuna varılmıştır.

Grzesiak et al. (2006), 305 günlük süt verimini tahmin etmek için yapay sinir ağları ve Wood metotlarını uygulamışlardır. Metotların performanslarını karşılaştırmak için nispi yaklaşık hatalar (NYH) ve hata kareler ortalamasının karekökü (HKOK) ve belirleme katsayısı  $R^2$  (%) gibi uyum ölçütleri baz alınmıştır. Çalışmanın sonucunda, Wood metodu yerine yapay sinir ağları kullanımının daha doğru olabileceği sonucuna varılmıştır.

Eyduran et al. (2008), Akkaraman ırkının varyeteleri olan Norduz (103 baş) ve Karakaş (101 baş) kuzuları üzerinde yaptıkları çalışmada, doğum ağırlığı üzerinde etkili olan faktörleri belirlemek amacıyla regresyon ağacı metodunu kullanmışlardır. CHAID algoritması ile oluşturulan regresyon ağacı diyagramına bakıldığında, doğum ağırlığını birinci dereceden doğum tipinin etkilediği görülmektedir (Düz.  $P < 0.01$ ). Buna ek olarak, doğum ağırlığını ikinci dereceden ırk ve cinsiyet etkilerken, üçüncü dereceden ise ana yaşının etkilediği belirlenmiştir (Düz.  $P < 0.01$ ).

Mendeş et al. (2009), 224 adet Ross-308 etçi piliçlerin kesim ağırlığını etkileyen değişkenlerin belirlenmesinde, regresyon ağacı analiz metodunu kullanmışlardır. Araştırmacılar, modele dahil edilen sekiz bağımsız değişkenden sadece dördünün (canlı



ağırlık, göğüs kemiği uzunluğu, bacak genişliği ve göğüs çevresi) kesim ağırlığı üzerinde etkili olduğu bildirmişlerdir.

Nisbet et al. (2009), SPSS, STATISTICA ve SAS paket programlarını kullanarak CART, CHAID, Boosted Classifiers ve Regression, MARS, Regresyon ve Sınıflama için Random Forest, Makine Öğrenme (Bayesyan, Destek Vektörleri, En yakın komşuluk, Bağımsız Bileşenler) gibi ileri veri madenciliği algoritmalarını detaylı olarak çalışmışlardır.

Bakır et al. (2010), farklı çiftliklerde yetiştirilen Holstein-Friesian ırkı sığırlar üzerinde yaptıkları çalışmada kuruda kalma süresi, laktasyon sırası, işletme, buzağılama mevsimi ve yaşı gibi çevre faktörlerinin 305 günlük süt verimi üzerindeki etkilerini regresyon ağacı metodunu kullanarak tespit etmişlerdir. Çalışmada kuruda kalma süresi (Düz<0.01) ve laktasyon sırasının (Düz<0.01), 305 günlük süt verimi üzerinde birinci derecede etkili faktörler olduğu işletme (Düz<0.01) ve buzağılama mevsimi (Düz<0.01) faktörlerinin ikinci derecede etkili faktörler olduğu ve buzağılama yaşı faktörünün (Düz<0.01) ise üçüncü derecede etkili olduğu tespit edilmiştir.

Ankaralı ve ark. (2010), yaptıkları çalışmada regresyon modellerinden birisi olan MARS modelinin özellikleri tanımlanarak diğer regresyon modellerine göre üstün olduğu durumları gösterilmiş ve uygulama adımları üzerinde çalışmışlar. MARS algoritmasının avantajı, bağımlı ve bağımsız değişkenlerin dağılımlarına ilişkin herhangi bir varsayım yoktur. Birden fazla bağımlı ve bağımsız değişkenlerin aynı anda analiz edilmesini sağladığını göstermişlerdir.

Zakeri et al. (2010), çocuk ve ergenlerde enerji harcamasının tahmini için çok değişkenli uyumlu regresyon uzanımları tekniği kullanmışlardır.

Tunay (2010), Türkiye'de muhtemel bankacılık krizlerini tahmin etmekte kullanılabilecek bir erken uyarı modeli geliştirmek için doğrusal ve parametrik olmayan bir tahminleme yöntemi olan MARS yöntemi kullanmıştır.

Acciani et al. (2011), gayrimenkul ve değerlendirme belirleyicileri arasındaki karmaşık desenleri ve etkileşimini ortaya koymak amacıyla MARS modellerinden faydalanmışlardır.

Grzesiak et al. (2011), yapay tohumlama açısından sıkıntılı olan hayvanların tanımlanmasında MARS modelini kullanmışlardır. MARS ile ilgili yapılan taramalarda ekonomi, bankacılık ve ticaret gibi alanlarda birtakım çalışmalara rastlanılmıştır.

Tunay (2011), Türkiye'de durgunlukları tahmin etmek amacıyla MARS yöntemi kullanmıştır. Araştırmacı 1986-2010 arası dönemi kapsayan üç aylık verileri kullanmıştır. Çalışmanın sonunda, Türkiye'deki durgunluklar bakımından MARS modelinin üstün performans gösterdiği belirlenmiştir.

Doğan ve Toprak, (2012), Diyarbakır Ticaret Borsası'nın 50 üyesiyle karşılıklı yapılan anket sonuçlarını değerlendirmek için MARS yöntemi uygulamıştır.

Grzesiak and Zaborski. (2012), hayvancılık alanında, CART (Classification and Regression Tree), CHAID (Chi-Square Automatic Interaction and Detection), MARS (Multivariate Adaptive and Regression Splines), ANN (Artificial Neural Network) ve diğer makine öğrenme metotları (Naive Bayes Classifier (NBC), Destek Vektör Makinaları (Support Vektör Machines =SVM), ve k-en yakın komşuluklar (k-nearest neighbors=k-NN) gibi veri madenciliği metotlarının teorik alt yapılarını açıklamışlardır.

Yakubu et al. (2012), yaptıkları çalışmada, Uda koçlarından alınan farklı bazı morfometrik ölçüler kullanarak canlı ağırlığı tahmin etmek için CART regresyon ağacı metodunu uygulamışlardır. Regresyon ağacı analizi sonuçları incelendiğinde, göğüs çevresi 94,05 cm' den geniş ve yüz uzunluğu 20,85 cm'den daha uzun olan hayvanların daha ağır olabileceği beklenmektedir. Sonuç olarak, araştırmacılar elde edilen bu bilgilerin, bakım-idare, seleksiyon ve ıslah çalışmaları açısından faydalı olabileceğini beyan etmişlerdir.

Mohammad et al. (2012), Pakistan'da bulunan farklı koyun ırkları (Mengali-48, Balochi-48, Harnai-48, Beverigh-47, Rakshani-48) üzerinde yürüttükleri bir çalışmada regresyon ağacı yöntemiyle, cidago yüksekliği, vucut uzunluğu ve göğüs çevresi gibi vücut ölçülerini kullanarak canlı ağırlık tahmini yapmışlardır. Tüm koyunlar içerisinde göğüs çevresi 89 cm'den büyük olan koyunların en yüksek canlı ağırlığa (48,6 kg) sahip olduğu belirlenmiştir. Regresyon ağacına göre canlı ağırlık üzerinde en etkili değişkenin, göğüs çevresi olduğu belirlenmiştir.

Eyduran et al. (2013a), 46 baş Türk Saanen keçilerinde laktasyon süt verimi, somatik hücre sayısı ve meme özellikleri arasındaki ilişkiyi, Exhaustive CHAID algoritması ile oluşturulan regresyon ağacı diyagramı ile incelemiştir. Ağaç diyagramı sonuçları incelendiğinde, en yüksek süt veriminin, meme açısı 50<sup>0</sup> dereceden geniş keçilerden elde edildiği bildirilmiştir. Meme açısı 50<sup>0</sup> dereceden daha dar olan keçiler içerisinde laktasyon süresi 191 günden uzun olanların süt veriminin, laktasyon süresi 191 ve 191 günden daha kısa olanlardan (192,25 lt) daha yüksek olduğu saptanmıştır (Düz. P<0.05). Meme açısı 50<sup>0</sup> dereceden dar olan keçilerin süt verimi, laktasyon süresinden etkilenmektedir (Düz. P<0.05). Ayrıca meme açısı 50<sup>0</sup> 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. (2013b), Mengali koyunlarının doğum ağırlığı üzerinde etkili olan çevre faktörlerini belirlemek için CHAID, algoritmasını kullanmışlardır. Çalışmada en etkili değişkenin doğum tipi olduğu belirlenmiştir. Bu araştırmada, 59 aydan daha yaşlı analardan doğan kuzuların en yüksek doğum ağırlığına sahip olduğu tespit edilmiştir. Ana yaşı faktörü, tekiz ve erkek kuzuların doğum ağırlığı üzerinde etkili olmuştur. Ancak, tekiz ve dişi kuzuların doğum ağırlığını etkileyen faktörün doğum yılı olduğu bildirilmiştir.

Khan et al. (2014), Harnai koyunlarına ait vücut ölçülerinden (vücut uzunluğu, cidago yüksekliği, göğüs çevresi, karın çevresi, yüz uzunluğu, kulaklar arası uzunluk, kuyruk genişliği ve uzunluğu) canlı ağırlığı tahmin etmek için Exhaustive CHAID algoritması kullanmışlardır. Sonuç olarak, canlı ağırlığı birinci derecede etkileyen en önemli bağımsız değişkenin yüz uzunluğu olduğu (Düz. P=0.000), ve bunu ikinci dereceden cidago yüksekliği (Düz. P=0.000), ve göğüs çevresi (Düz. P=0.000), üçüncü dereceden göğüs çevresi (Düz. P=0.000) ve vücut uzunluğu (Düz. P=0.001) değişkenlerinin izlediği belirlenmiştir. Böylece, bu algoritma ile canlı ağırlığa ait toplam varyasyonun % 84.4'ünün (R<sup>2</sup>) açıklandığı saptanmıştır.

Abayomi-Alli et al. (2015), öngörü yüz resim doğrulama ve kalite değerlendirme ile ilgili MARS modellerinden faydalanmaya çalışmışlardır.

Çetin ve Mikail (2016), yapmış oldukları derleme çalışmasında; k ortalamaları yaklaşımı, k-en yakın komşu yaklaşımı, çok değişkenli uyarlanırlı regresyon eğrileri (Multivariate Adaptive Splines, MARS), Bayes sınıflandırıcıları (Naive Bayesian Classifiers, NBC), yapay sinir ağları (Artificial Neural Networks, ANN), destek vektör makineleri (Support Vector Machines, SVM), karar ağaçları gibi çok yararlı veri madenciliği algoritmaları konusunda teorik bilgiler vermişlerdir.

Adamczyk et al. (2016), yapay sinir ağları ve boosted sınıflama ağaçlarını ayıklama nedenlerini tanımlamak için kullanmışlardır. Hayvancılık alanında sınıflama yapma amacıyla MARS modeli kullanılmıştır..

Eyduran, (2016), yapmış olduğu çalışmada regresyon analizinde kullanılan bazı varsayımların ihlal edilmesi durumunda veri madenciliği algoritmalarının kolaylıkla kullanılabilirdiğinden bahsetmiştir. Bu araştırmacı, hayvancılık alanında az uygulamaya sahip olan MARS modelinin bağımsız değişkenler arasındaki yüksek dereceli intereaksiyonları ortaya koymada yararlı bir analiz olduğunu vurgulamışlardır.

Karadaş et al. (2017), 2 yaşlı 304 baş Mengali koyunu üzerinde yaptıkları çalışmada yedi morfolojik özelliği (cidago yüksekliği, vücut uzunluğu, göğüs çevresi, alın uzunluğu, alın genişliği, kuyruk uzunluğu ve kulak uzunluğu) kullanarak canlı ağırlık özelliğini tahmin etmeye çalışmışlardır. Araştırmacılar, MARS veri madenciliği algoritmasının CHAID algoritmasına nazaran daha iyi sonuç verdiğini tespit etmiştir.

Eyduran et al. (2017), Mengali koçlarından bazı vücut özellikleri ile kullanarak canlı ağırlığı tahmin etmek amacıyla MARS yöntemini kullanmışlardır. Non-parametrik teknik olan MARS modelinin canlı ağırlığı tahmini bakımından iyi sonuçlar verdiğini belirtmişlerdir.

### **3. MATERYAL ve METOT**

#### **3.1. Materyal**

MARS algoritmasını uygulamak amacıyla bu tez çalışmasında kullanılan veri seti Mohammad et al. (2012)'den alınmıştır.

Canlı ağırlıklarını tahmin etmek amacıyla yaşları 11-13 ay arasında değişen Balochi (48), Harnai (48), Beverigh (47) ve Rakhshani (48) olmak üzere toplam 191 baş erkek toklunun cidago yüksekliği, vücut uzunluğu ve göğüs çevresi gibi vücut ölçütleri alınmıştır.

#### **3.2. Metot**

Genel anlamda regresyon analizi bir bağımlı değişken ile bir ya da birden fazla bağımsız değişken arasındaki ilişkinin ortaya konulduğu analiz tekniğidir. Regresyon analizinde amaç; kolay ölçülebilen özellikler yani değişkenlerden yola çıkarak zor ölçülebilen bağımlı değişken için tahmin denklemi geliştirmektir. Regresyon analizinde temel amaç; tahmin hatasını azaltan bağımsız değişkenleri tespit etmektir. Diğer bir ifadeyle; bağımlı değişkenin gerçek değeri ile (regresyon analizi ile tahmin edilen) tahmin değeri arasındaki korelasyonu arttırmak ya da farkı azaltmaktır. Çoklu Doğrusal Regresyon Analizine (Multiple Linear Regression Analysis) alternatif analiz teknikleri bulunmaktadır. Bunlar;

1. Ridge Regresyon Analizi (Ridge Regression Analysis)
  2. Adımsal Regresyon Analizi (Stepwise Regression Analysis)
  3. Faktör Analizi Skorları ile Çoklu Regresyon Analizi (Factor Analysis Scores in Multiple Linear Regression Analysis)
  4. Temel Bileşenler Analizi Skorları ile Çoklu Regresyon Analizi (Principal Component Analysis Scores in Multiple Linear Regression Analysis) şeklindedir. Bununla birlikte çoklu regresyon analize alternatif olarak veri madenciliği algoritmaları da kullanılabilir. Bunlar;
- ✓ CART algoritması (Classification and Regression Tree)

- ✓ CHAID algoritması (Chi-Square Automatic Interaction Detector)
- ✓ Exhaustive CHAID algoritması(Exhaustive Chi-Square Automatic Interaction Detector)
- ✓ Çok katmanlı Perceptron (Multilayer Perceptron Neural Network Algorithm) ve
- ✓ Radyal Temel Fonsiyonlu Yapay sinir ağı algoritması (Radial Basis Function Neural Network Algorithm) şeklinde sıralanabilir.

Regresyon analizi kapsamında CHAID, CART, Exhaustive CHAID, MLP ve RBF gibi algoritmalar nominal, ordinal ve sürekli deęişkenler için kullanılabilir. CHAID, CART ve Exhaustive CHAID algoritmaları bir bağımlı ve birden fazla bağımsız deęişken yapısı için rahatça kullanılabilir. Ancak literatürde regresyon analizi kapsamında MARS modelinin hayvancılık alanında kullanımı yetersizdir.

MARS algoritması Stanford'da fizikçi ve istatistikçi olan Jerome FRIEDMAN tarafından 1991 yılında geliştirilmiştir. Bu algoritma; bağımlı ve bağımsız deęişkenlerin dağılımlarına ilişkin herhangi bir varsayımının olmaması ve birden fazla bağımlı ve bağımsız deęişkenlerin aynı anda analiz edilebilmesi gibi avantajlara sahiptir.

CHAID, CART ve Exhaustive CHAID algoritmaları tahmin denklemi yerine ağaç yapısı şeklinde görsel bir karar ağacı diyagramı oluştururken, MARS algoritması az sayıda bağımsız deęişken ile tahmin doğruluęu yüksek tahmin denklemi oluşturur. Bağımlı ve bağımsız deęişken arasında doğrusal olmayan yapıdaki ilişkileri (ölçüm deęerlerine en yakın yakın olan) doğrusal yapıya dönüştürmeye imkan vererek tahmin hatasının azaltılmasını sağlar.

Parçalı temel fonksiyonlar ve bunların bileşimlerini kullanarak ve regresyon modellerindeki hem ileri doğru hem de geriye doğru gelişme algoritmalarından faydalanarak MARS modeli oluşturulabilir. Model oluşumu iki aşamada şekillenir.

Birinci aşama: Olası bütün temel fonksiyonlar şekillendirilir. Çok büyük bir model elde edilinceye kadar yani modelin komplikasyonu en üst seviyeye ulaşmaya kadar ilave edilen temel fonksiyonlarla model gelişir.

İkinci aşama: Temel fonksiyonlar tahmin hatası minimum olacak şekilde modelden çıkarılır. Bu aşamaya budama (prunning) aşaması denir. Bu aşamayı oluşturabilmek için MARS bir geriye doğru adım algoritması uygulanır (Ankaralı ve ark., 2010).

Oluşturulan en üst düzeydeki model budanarak yani önemli bağımsız değişkenlerin ve bu değişkenlerin etkileşimi belirlenerek, hata kareler toplamı minimum olan en uygun model biçimlendirilir. Budama algoritması en genel geçer olarak olağan bir şekilde Genelleştirilmiş Çapraz Geçerlilik (Generalized Cross Validation, GCV) tekniği ile gerçekleştirilir.

MARS modeli (3.1) nolu eşitlikteki gibi tanımlanabilir.

$$Y = \beta_0 + \sum_{k=1}^K \beta_k B_k(X) \dots\dots\dots (3.1)$$

Burada;

k:Düğüm sayısı

K:Temel fonksiyon sayısı

X:Bağımsız değişken

$B_k$ :Modeldeki sabit sayısı

$\beta_0$ :Regresyon katsayısı

$B_k(x)$ :k.Temel fonksiyon biçimindedir.

Temel fonksiyon ise aşağıdaki gibi tanımlanır

$$B_k(X) = \prod_{j=1}^{j_k} [S_{kj}(X_{wkj} - t_{kj})] \dots\dots\dots (3.2)$$

Buradaki;

$J_k$  interaksiyon derecesini göstermektedir.

$[.]_+ = \max[0, .]_+$

$s_{kj} : \in [\pm 1]$

$t_{kj}$ : Dügüm deęeri ve

$X_{wkj}$ : Baęımsız deęişken deęerini göstermektedir.

Tahmin hatasını en küçük yapan en uygun ölçümle her deęişken ve mümkün bütün düğümler için temel fonksiyon uzayı araştırılır. Maksimum karmaşıklıęa ulaşan bir modeli oluşturuncaya kadar bütün düğümlere ve deęişkenlere bakma işlemi tekrarlanır. Son olarak, budama prosedürüne başvurulur. Oluşturulan maksimum model budanarak yani önemli baęımsız deęişkenler ve bu deęişkenlerin interaksiyonları belirlenerek, hata kareler toplamı en küçük olan en uygun model oluşturulur. Budama algoritması en yaygın olarak genişletilmiş çapraz geçerlilik (Cross Validation, GCV) teknięi ile yapılır. Genelleştirilmiş çapraz geçerlilik hem artıkların hatasını, hem de model karmaşasını hesaba katar ve (3.3) nolu eşitlik yardımıyla hesaplanır.

$$GCV = \frac{\sum_{i=1}^N (y_i - f(x_i))^2}{\left[1 - \frac{C}{N}\right]^2} \dots\dots\dots(3.3)$$

Burada,

**N:** Veri setindeki toplam hayvan sayısını,

**d:** Etkili serbestlik derecesi göstermektedir.

Etkili serbestlik derecesi baęımsız temel fonksiyonların sayısına eşittir.

**C:** eklenen temel fonksiyonların maliyet – karmaşıklık (cost-complexity) ölçüsüdür.



Hesaplamalar şunu göstermiştir ki en iyi C değeri için  $2 < d < 3$  bulunmuştur. Ancak, analizler R yazılımı ile yapılmış olup, GCV ölçütü penalty= -1 kısıtlamasından dolayı aşağıdaki 4 nolu eşitliğe dönüştürülmüştür.

$$GCV = \frac{\sum_{i=1}^N (y_i - f(x_i))^2}{N} \quad \dots\dots\dots(3.4)$$

#### Model uyum iyiliği ölçütleri

Bu çalışmada oluşturulan MARS modellerinin tahminleme performanslarını değerlendirmek için aşağıdaki uyum iyiliği testlerinden yararlanılmıştır (Koc et al. 2017). Uyum iyiliğine ait formüller aşağıda verilmiştir.

##### 1. Belirleme katsayısı

$$R^2 = \left[ 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right] \quad \dots\dots\dots (3.5)$$

##### 2. Düzeltilmiş Belirleme Katsayısı

$$\bar{R}^2 = \left[ 1 - \frac{\frac{1}{n-k-1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2} \right] \quad \dots\dots\dots (3.6)$$

##### 3. Standart sapma oranı

$$SS_{oran} = \sqrt{\frac{\frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}} \dots\dots\dots (3.7)$$

Burada;

$$\varepsilon_i = Y_i - \hat{Y}_i \text{ (Gerçek Canlı Ağırlık– Tahmin edilen Canlı Ağırlık)}$$

n= Çalışmada kullanılan toplam hayvan sayısını

$Y_i$ : i. toklunun canlı ağırlığını (kg)

$\hat{Y}_i$ : i. toklunun tahmin edilen canlı ağırlığını (kg)

$\bar{Y}$ : Ortalama toklu ağırlığını,

$\varepsilon_i$ : i. tokluya ait hata değerini,

$\bar{\varepsilon}$ : Ortalama hata değerini,

k: Modeldeki terim sayısını ifade etmektedir.

#### 4. BULGULAR ve TARTIŞMA

Canlı ağırlığı tahmin etmek amacıyla maksimum interaksiyon derecesi 2 için MARS algoritmasıyla oluşturulan tahmin denklemine ilişkin katsayılar ve temel fonksiyonlar Çizelge 4.1'de verilmiştir. Budamadan önce maksimum terim sayısı (nk) R programında 35 olarak ayarlanmış, ancak bununun 26'sı MARS tahmin denklemine dahil edilmiştir. Ayrıca, Çizelge 4.1'de elde edilen sonuçlar  $\text{penalty}=-1$  ile elde edildiğinde dolayı genelleştirilmiş çapraz geçerlilik  $\text{GCV}=\text{RSS}/n(115.8274/191)$  şeklinde rahatça hesaplanabilmektedir. Genelleştirilmiş çapraz geçerlilik 0.606426 olarak tahmin edilmiştir. Burada nfold yani çapraz geçerlilik 10 olarak ayarlandığında son çapraz geçerlilik  $R^2(\text{CVRsq})$  değeri 0.922 gibi yüksek bir değer elde edilmiştir. MARS algoritması için genelleştirilmiş çapraz geçerlilik değerinin en düşük olduğu modelin en iyi model olduğu düşünülür. Analiz sonucunda  $\text{GRsq}$  (modelin tahminleme gücü) ve  $R^2(\text{CVRsq})$  değerleri yaklaşık 0.96 gibi oldukça yüksek bir değer olarak tahmin edilmiştir. Ayrıca standart sapma oranı 0.02052 değeri 0.10 dan oldukça düşük olduğu için modelin mükemmel bir uyuma sahip olduğu kanıtlanmıştır.

Diğer veri madenciliği algoritmalarında olduğu gibi MARS algoritmasında da temel amaç, bağımlı değişkene (canlı ağırlık) ait gözlenen ve tahmin edilen değerler arasındaki farkın azaltılmasıdır. Diğer bir ifadeyle, iyi bir algoritmada söz konusu bağımlı değişkene ait gözlenen ve tahmin edilen değerler arasındaki Pearson korelasyonunun oldukça yüksek olması istenir. Çizelge 4.1 incelendiğinde, CY ve GC bağımsız değişkenlerinin CA üzerindeki etkisinin ırka göre değiştiği söylenebilir. Ayrıca CY ile VU değişkenlerinin CA üzerindeki etkisi birbirine bağlıdır. VU değişkeninin CA üzerindeki etkisi GC'ne bağlı olarak değişebilmektedir. Örneğin,  $\text{VU}>88.9$  cm ve  $\text{GC}>71.12$  cm bir koyunda, modeldeki  $-0.3146535 \cdot h(\text{VU}-88.9) \cdot h(\text{GC}-71.12)$  terimi CA bağımlı değişkenini azaltıcı bir etkiye sahip olur. GCV ve RSS kriterleri baz alındığında bağımsız değişkenlerin relatif önem dereceleri aşağıdaki gibi sıralanır.

$\text{GC}(100) > \text{VU}(76.7) > \text{CY}(65.6) > \text{Rakhashni}(47.7) > \text{Harnai}(38.9) > \text{Beverigh}(0)$

Bununla ilgili R program komutları aşağıda verilmiştir.

```
>d=read.table("C:/analiz.txt", header=T)
```

```

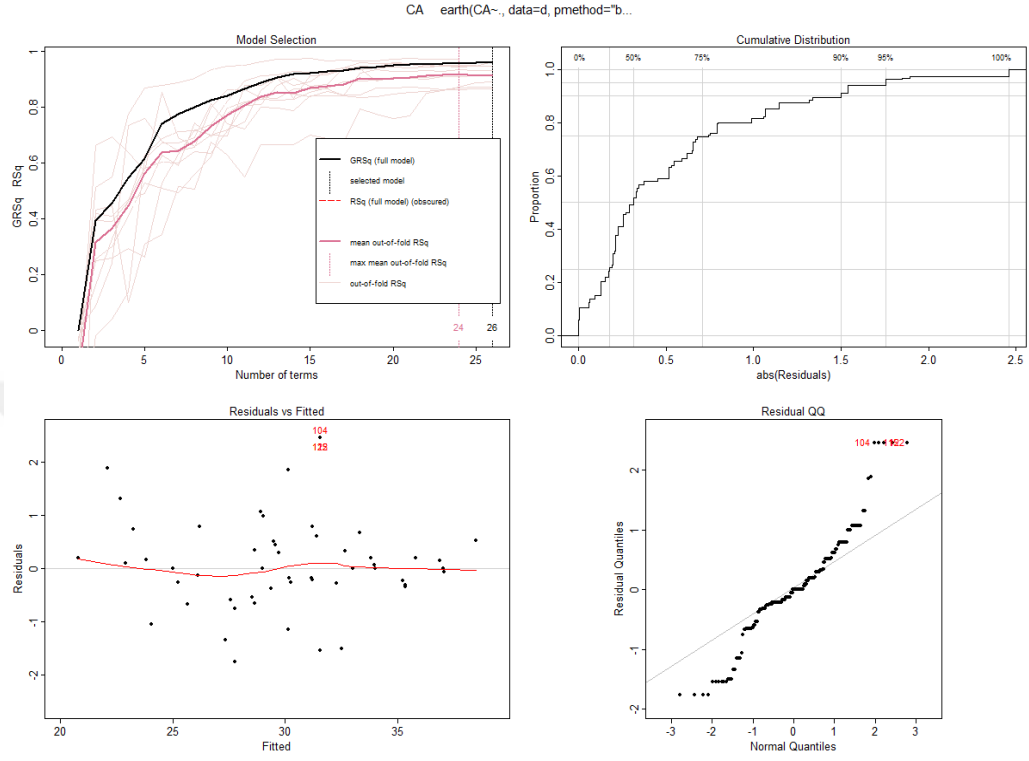
>install.packages("earth")
> library(earth)
> mars2=earth(formula=CA~.,data=d, nfold=10, ncross=20, penalty=-1,degree = 2,
nk=35, pmethod="backward", keepxy = T)
> summary(mars2)

```

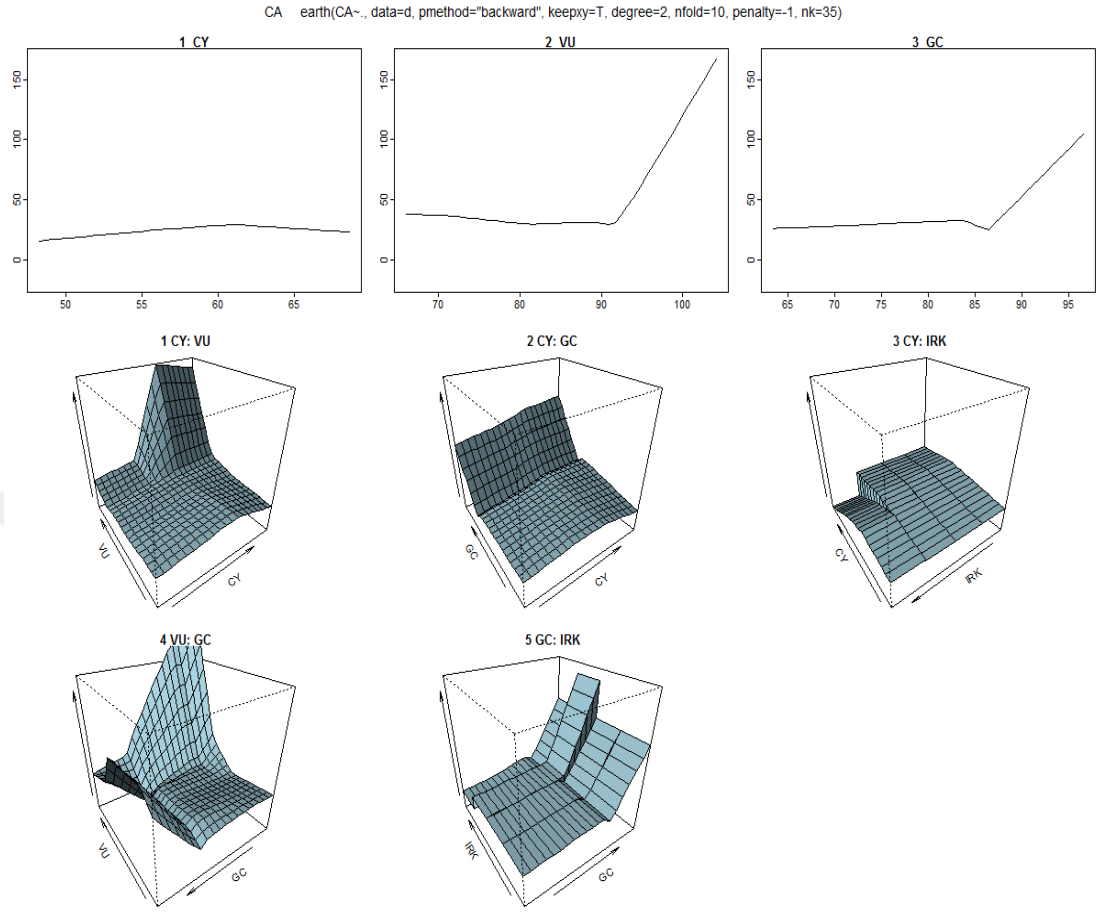
**Çizelge 4.1. MARS tahmin denklemine ait katsayılar ve terimler**

Terimler	Katsayılar
(sbt)	25.1005278
h(55.88-CY)	-1.1781510
h(CY-55.88)	0.8152376
h(CY-60.96)	-0.5304742
h(81.28-VU)	0.6333313
h(VU-81.28)	0.3992706
h(VU-93.98)	2.3580003
h(71.12-GC)	0.2260831
h(GC-71.12)	3.3437537
h(GC-83.82)	-3.7540160
h(GC-86.36)	11.2471213
h(CY-55.88) * IRKRakhashni	-3.4846677
h(GC-83.82) * IRKHarnai	3.9439179
h(GC-71.12) * IRKRakhashni	0.7498772
h(CY-55.88) * h(VU-91.44)	1.9678530
h(CY-55.88) * h(91.44-VU)	0.0092132
h(CY-60.96) * h(VU-91.44)	-2.1200919
h(CY-60.96) * h(91.44-VU)	-0.1189799
h(CY-55.88) * h(GC-73.66)	0.0756330
h(CY-55.88) * h(73.66-GC)	-0.0981755
h(VU-71.12) * h(GC-71.12)	-0.1961108
h(VU-81.28) * h(GC-68.58)	0.0067209
h(VU-81.28) * h(68.58-GC)	0.5644906
h(VU-83.82) * h(GC-71.12)	-0.0599798
h(88.9-VU) * h(GC-71.12)	-0.1809336
h(VU-88.9) * h(GC-71.12)	-0.3146535

GCV 0.606426 RSS 115.8274 GRSq 0.9578761 RSq 0.9578761 CVRSq 0.9218641

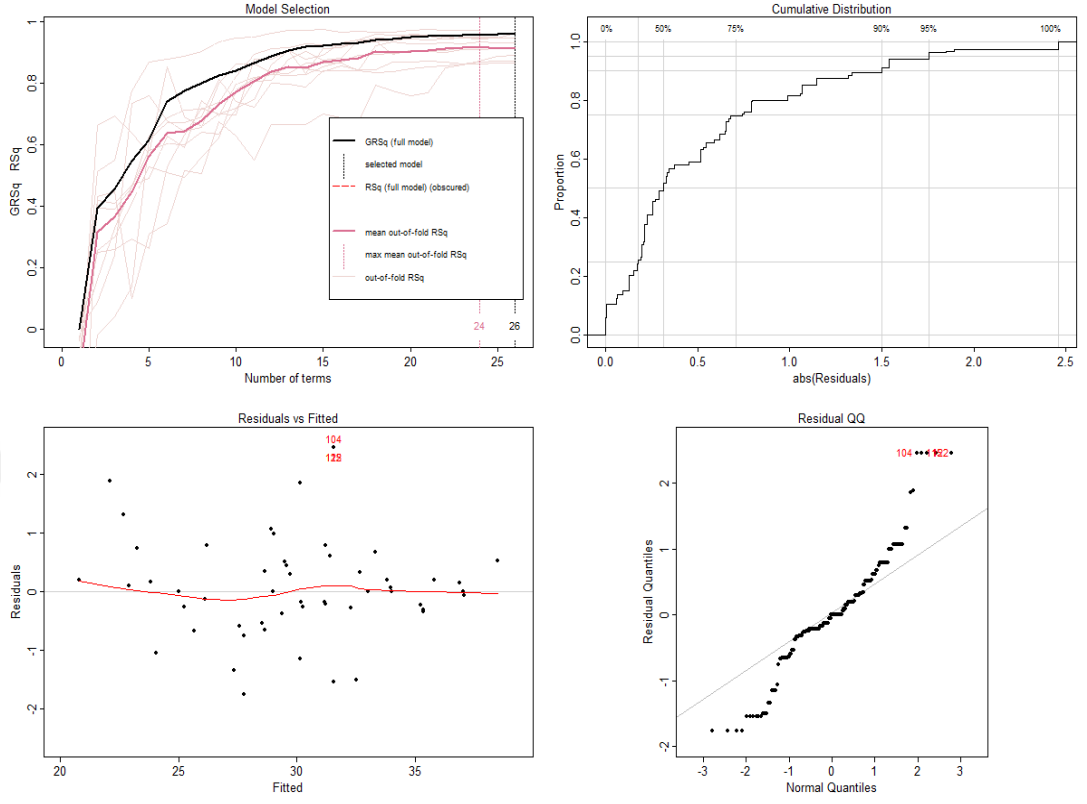


**Şekil 4.1.**Seçilen modele ilişkin belirlenme katsayıları ve hata terimlerine ait grafikler

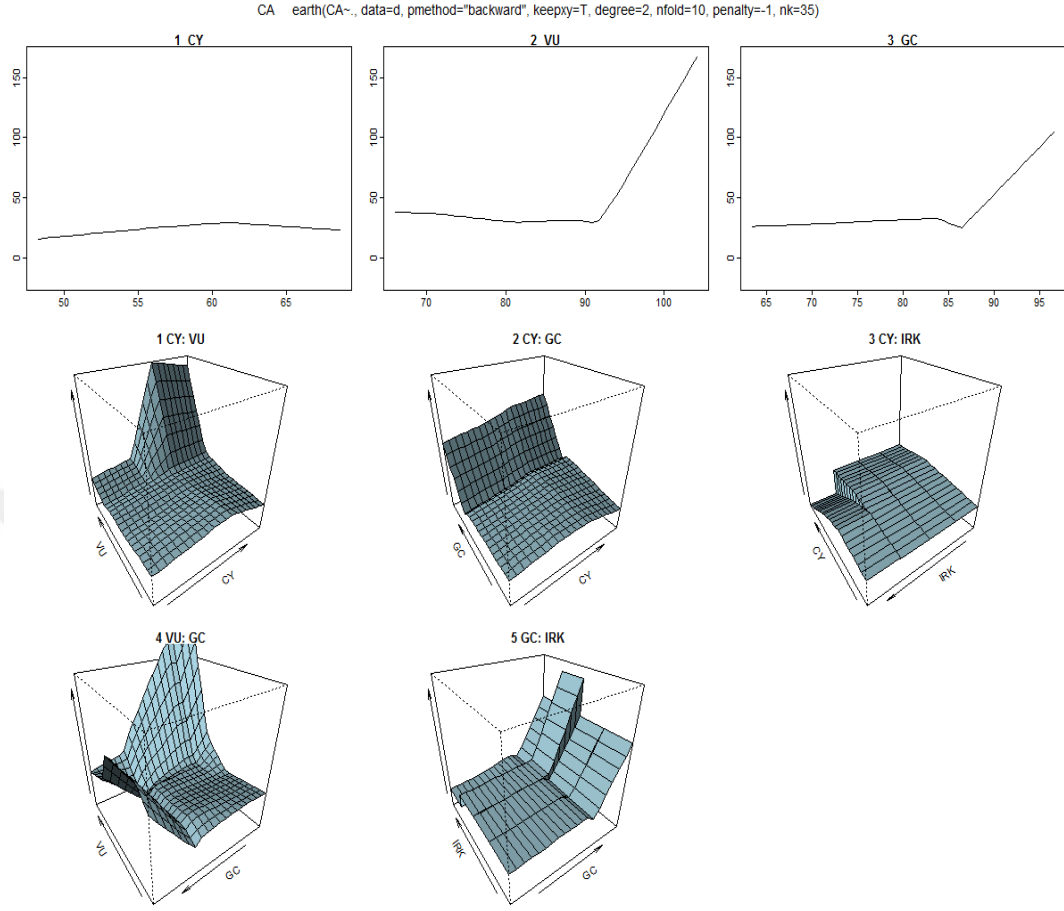


Şekil 4.2. Bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiler.

CA earth(CA--, data=d, pmethod="b...



Şekil 4.3. Seçilen modele ilişkin belirlenme katsayıları ve hata terimlerine ait grafikler



**Şekil 4.4.** Bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiler

Canlı ağırlığı tahmin etmek amacıyla maksimum interaksiyon derecesi 2 için MARS algoritmasıyla oluşturulan tahmin denkleminin katsayıları ve temel fonksiyonları Çizelge 4.2'de verilmiştir. Budamadan önce maksimum terim sayısı (nk) R programında 38 olarak ayarlanmış, tamamı MARS tahmin denkleminde dahil edilmiştir. Aynı zamanda, Çizelge 4.2'de elde edilen sonuçlar  $\text{penalty}=-1$  ile elde edildiğinde dolayı genelleştirilmiş çapraz geçerlilik  $\text{GCV}=\text{RSS}/n$  ( $55,99683/191$ ) şeklinde kolayca hesaplanabilmektedir. Genelleştirilmiş çapraz geçerlilik 0,2931771 olarak tahmin edilmiştir. Burada nfold yani çapraz geçerlilik 10 olarak ayarlandığında son çapraz geçerlilik  $R^2(\text{CVRsq})$  değeri 0,945 gibi yüksek bir değer olarak bulunur. MARS algoritması için genelleştirilmiş çapraz geçerlilik değerinin en düşük olduğu modelin en iyi model olduğu düşünülür. Analiz sonucunda  $\text{GRsq}$  (modelin tahminleme gücü) ve  $R^2(\text{CVRsq})$  değerleri yaklaşık 0,97 gibi oldukça yüksek bir değer olarak tahmin



edilmiştir. Ayrıca standart sapma oranı 0.01427 değeri 0.10'dan oldukça düşük olduğu için modelin mükemmel bir uyuma sahip olduğu kanıtlanmıştır.

Diğer veri madenciliği algoritmalarında olduğu gibi MARS algoritmasında da asıl amaç, bağımlı değişkene (canlı ağırlık) ait gözlenen ve tahmin edilen değerler arasındaki farkın azaltılmasıdır. Diğer bir ifadeyle, iyi bir algoritmada söz konusu bağımlı değişkene ait gözlenen ve tahmin edilen değerler arasındaki Pearson korelasyonunun oldukça yüksek olması istenir. Çizelge 4.2 incelendiğinde, CY ve GC bağımsız değişkenlerinin CA üzerindeki etkisinin ırka göre değiştiği söylenebilir. Aynı zamanda CY ile VU değişkenlerinin CA üzerindeki etkisi birbirine bağlıdır. VU değişkeninin CA üzerindeki etkisi GC'ne bağlı olarak değişebilmektedir. Örneğin,  $VU > 81.28$  cm ve  $GC > 73.66$  cm olan bir koyunda, modeldeki  $0.0093125 * h(VU - 81.28) * h(GC - 73.66)$  terimi CA bağımlı değişkenini artırıcı bir etkiye sahip olur. İkinci model için GCV ve RSS kriterleri baz alındığında bağımsız değişkenlerin relatif önem dereceleri aşağıdaki gibi sıralanır.

GC(100) > VU(77.3) > CY(73.1) > Rakhashni(56.1) > Harnai(37.3) > Beverigh(15.5)

Bununla ilgili R program komutları aşağıda verilmiştir.

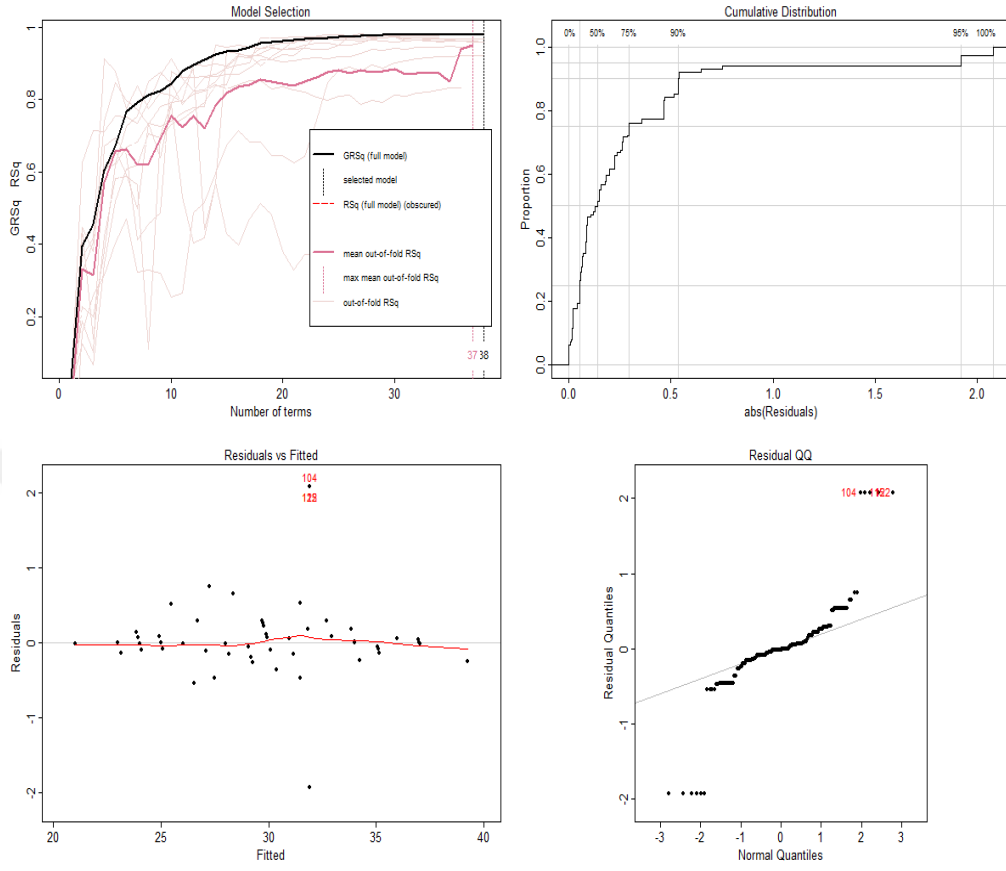
```
>d=read.table("C:/analiz.txt", header=T)
>install.packages("earth")
> library(earth)
>mars2=earth(formula=CA~.,data=d, nfold=10, ncross=20, penalty=-1,degree = 2, nk=
150, pmethod="backward", keepxy = T)
> summary(mars2)
> plot(mars2)
> plotmo(mars2)
```

**Çizelge 4.2.** MARS tahmin denklemine ait katsayılar ve terimler

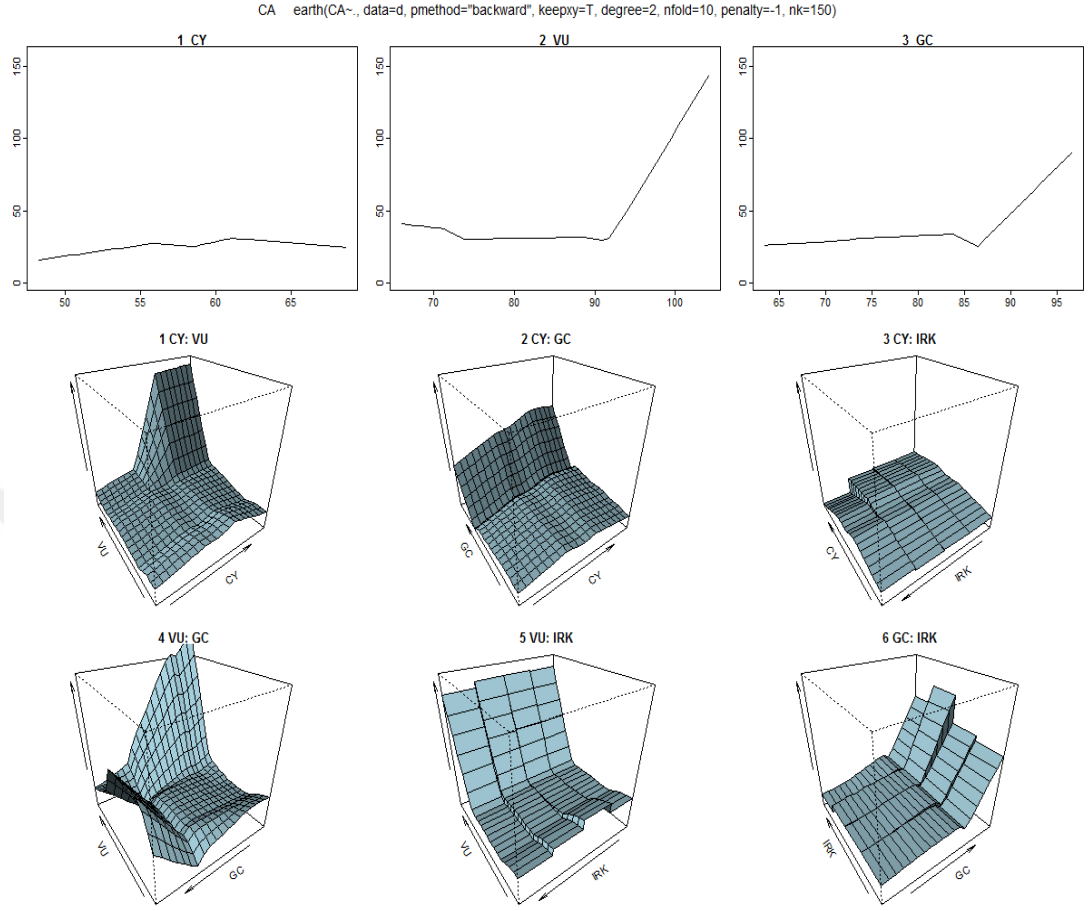
Terimler	Katsayılar
(sbt)	17.6127739
h(55.88-CY)	-1.3363471
h(CY-55.88)	-1.6327272
h(CY-58.42)	3.2580405
h(CY-60.96)	-1.4797150
h(VU-71.12)	-1.1931640
h(VU-73.66)	2.8552556
h(81.28-VU)	1.4876924
h(VU-81.28)	-1.2963972
h(VU-93.98)	1.3828350
h(71.12-GC)	0.0310414
h(GC-71.12)	6.6538548
h(GC-83.82)	-3.7845641
h(GC-86.36)	9.8848615
h(CY-55.88) * IRKRakhashni	-1.9217400
h(81.28-VU) * IRKBeverigh	0.6846941
h(GC-71.12) * IRKBeverigh	0.2559542
h(GC-83.82) * IRKHarnai	3.0579932
h(GC-71.12) * IRKRakhashni	1.0238533
h(CY-55.88) * h(VU-91.44)	1.7303658
h(CY-55.88) * h(91.44-VU)	0.0488117
h(CY-60.96) * h(VU-91.44)	-1.7124261
h(CY-60.96) * h(91.44-VU)	-0.1516419
h(CY-55.88) * h(GC-73.66)	-0.0106929
h(CY-55.88) * h(73.66-GC)	-0.0765804
h(CY-60.96) * h(GC-76.2)	0.1077249
h(CY-60.96) * h(76.2-GC)	0.0498865
h(63.5-CY) * h(GC-71.12)	-0.0670504
h(CY-63.5) * h(GC-71.12)	-0.0587913
h(VU-71.12) * h(GC-71.12)	-0.3686860
h(81.28-VU) * h(GC-68.58)	-0.0510151
h(81.28-VU) * h(68.58-GC)	-0.0111393
h(VU-81.28) * h(GC-68.58)	0.0394031
h(VU-81.28) * h(68.58-GC)	0.2155141
h(VU-81.28) * h(GC-73.66)	0.0093125
h(VU-83.82) * h(GC-71.12)	-0.0408482
h(88.9-VU) * h(GC-71.12)	-0.3124303
h(VU-88.9) * h(GC-71.12)	-0.1346043

GCV 0.2931771 RSS 55.99683 GRSq 0.9796352 RSq 0.9796352 CVRSq 0.8753893

CA earth(CA~, data=d, pmethod="b...



Şekil 4.2.1. Seçilen modele ilişkin belirlenme katsayıları ve hata terimlerine ait grafikler



**Şekil 4.2.2.** Bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiler.

Literatüre bakıldığında MARS modeli ile canlı ağırlık tahmininin yapılmasıyla ilgili çok az sayıda çalışmalara rastlanılmıştır. Eyduran et al. (2017), bir yaşlı Mengali koçlarında bazı morfolojik ölçülerden canlı ağırlık tahmini etmek için MARS veri madenciliği algoritması kullanmış ve aşağıdaki tahmin denklemini geliştirmiştir.

$$CA = 34.53158 + 0.57169 * \max(0; GC - 83.82) - 0.54473 * \max(0; 83.82 - GC) - 0.01059 * \max(0; VU - 68.58) * \max(0; GC - 83.82)$$

Bu önceki çalışmada elde edilen uyum iyiliği ölçütlerine nazaran bu çalışmada elde edilen ölçütler daha iyi sonuçlar vermiştir.

Karadas et al. (2017) iki yaşlı mengali koyunlarında yaptığı çalışmada, vücut ölçülerinde canlı ağırlık tahmini yapmak amacıyla MARS algoritması kullanmıştır. Bu

çalışmada bir yaşlı koyunlardan elde edilen sonuçlar, Karadas et al. (2017)'nin sonuçlarından daha iyi bulunmuştur.

Bu çalışmada elde edilen uyum iyiliği ölçütlerine ait sonuçlar, farklı istatistik metotları kullanılan diğer çalışma sonuçlarından daha iyi bulunmuştur (Afolayan et al. 2006; Eyduran et al., 2009; Yakubu, 2012; Mohammad et al., 2012; Khan et al., 2014; Ali et al., 2015).

Literatürdeki söz konusu farklılıklar, hayvanın yaşı, ırkı, cinsiyeti, yetiştirme ve besleme koşulları, bağımsız değişkenlerin ve interaksiyon düzeylerinin farklı olmasına ve özellikle uygulanan istatistiksel metotların farklı olmasına atfedilebilir.

Canlı ağırlığı tahmin etmek amacıyla çoklu regresyon analiz yöntemi kullanılmış, elde edilen varyans analizi sonuçları çizelge 4.3' de verilmiştir.

**Çizelge 4.3.**Çoklu regresyon analizine ait varyans analizi tablosu

	Sd	KT	KO	F	Pr(>F)
CY	1	350.46	350.46	67.573	3.551e-14
VU	1	79.06	79.06	15.244	0.0001325
GC	1	856.43	856.43	165.132	< 2.2e-16
BREED	3	509.44	169.81	32.743	< 2.2e-16
HATA	184	954.29	5.19		

Yapılan çoklu regresyon analizinde modele dahil edilen tüm değişkenler canlı ağırlık üzerinde önemli etkiye sahip olmuştur.( $p < 0.01$ ) Canlı ağırlığa ilişkin toplam varyasyonun yaklaşık 2/3'ü açıklanmıştır. Çoklu regresyon analizine ait belirleme katsayısı 0.6529 , düzeltilmiş belirleme katsayısı 0.6416 ve standart sapma oranı ise 0.58911 olarak bulunmuştur. Uyum iyiliği ölçütleri bakımından MARS algoritması ile kurulan modellerin tahmin performansı, çoklu doğrusal modelden çok daha iyi bulunmuştur.

## 5. SONUÇ VE ÖNERİLER

Bu çalışmada, literatürdeki diğer çalışmalardan farklı olarak koyunlarda bazı morfolojik özelliklerden canlı ağırlığını tahmin etmek amacıyla MARS algoritması kullanarak yeni tahmin denklemleri geliştirilmiştir. MARS uygulamalarının tahmin performanslarının değerlendirilmesi için uyum iyiliği ölçütleri (GCV, RSS,  $GR^2$ ,  $R^2$ , Düzeltilmiş  $R^2$  ve Çapraz geçerlilik  $R^2$  ve gerçek ve tahmin değerleri arasındaki korelasyon katsayısı) kullanılmıştır. Burada;

- a) 26 terim içeren I. MARS tahmin modeli için canlı ağırlık bakımından gerçek ve tahmin edilen değerler arasındaki korelasyon katsayısı oldukça önemli ve yüksek bulunmuştur ( $r=0.980$ ,  $P<2.2e-16$ ). Ayrıca en düşük GCV (0.610) ve RSS (116) baz alınarak hesaplanan  $GR^2$  (0.960),  $R^2$  (0.960), düzeltilmiş  $R^2$  (0.954) ve çapraz geçerlilik  $R^2$  (0.922) uyum iyiliği ölçütlerinin oldukça iyi sonuç verdiği gözlenmiştir. Bu durum kurulan MARS modelinin canlı ağırlığını tahmin etmede oldukça doğru sonuçlar verdiğini belirlenmiştir.
- b) 38 terim içeren II. MARS tahmin modeli için canlı ağırlık bakımından gerçek ve tahmin edilen değerler arasındaki korelasyon katsayısı oldukça önemli ve yüksek olduğu belirlenmiştir ( $r=0.990$ ,  $P<2.2e-16$ ). Ayrıca en düşük GCV (0.293) ve RSS (56) baz alınarak hesaplanan  $GR^2$  (0.980),  $R^2$  (0.980), düzeltilmiş  $R^2$  (0.975) ve çapraz geçerlilik  $R^2$  (0.88) uyum iyiliği ölçütlerinin oldukça iyi sonuç verdiği gözlenmiştir. Bu durum II. MARS modelinin de canlı ağırlığını tahmin etmede oldukça güvenilir sonuçlar verdiğini belirlenmiştir.
- c) Her iki MARS uygulamasında her koyun ırkı için tahmin denklemi elde edilebilir.

En küçük GCV değerini üreten interaksiyon dereceleri ve uygun terim sayısının atanması iyi bir MARS modellemesinin yapılmasında önemlidir. Ayrıca, R programında MARS algoritmasında ceza (penalty) değerinin 2, 3 ve 4 yerine -1 olarak özelleştirilmesi daha güvenilir bir tahminleme yapılmasını olanak sağlayabilmektedir. Bununla beraber, en düşük GCV ve RSS değerini ile birlikte  $GR^2$ ,  $R^2$ , düzeltilmiş  $R^2$  ve çapraz geçerlilik  $R^2$  uyum iyiliği ölçütlerinin hepsinin yüksek olması önerilen bir

durumdur. Yüksek interaksiyonlu MARS algoritma uygulamalarında iyi bir sonuç elde etmek açısından terim sayısının ilk etapta yüksek olarak özelleştirilmesi önerilebilir. Mars algoritmalarında interaksiyon derecesi kademeli olarak artırılarak uyum iyiliği ölçütlerinin performansı her aşamada değerlendirilmelidir.

Çapraz geçerlilik değeri genelde literatürde 10 olarak alınmaktadır. Ancak bazı durumlarda, uyum iyiliği ölçütleri için optimum değerler 3 ya da 5 çapraz geçerlilik değerleri için elde edilebilir. Araştırmacıların daha iyi sonuçlar elde edebilmesi, bağımlı değişken üzerinde etkili olan bağımsız değişkenlerin seçimine bağlıdır. Bu nedenle, mevcut çalışmada canlı ağırlık üzerinde etkili olan morfolojik özellikler kullanılmıştır.

Bağımsız değişkenlere ait doğrusal olmayan ve interaksiyon etkilerinin ortaya konulması bakımından MARS algoritmasının kullanımı önerilebilir. Genel doğrusal modellere alternatif bir metod olarak düşünülebilir. MARS modelinin en önemli üstünlüğü birden fazla bağımlı ve bağımsız değişkenler arasındaki karmaşık ilişkileri ortaya koyabilmesidir. Bu bakımdan yapay sinir ağları algoritmalarına alternatif olarak kullanılabilir. Yanıt yüzeyi yönteminden (Response Surface Method) elde edilen verilere MARS modeli uygulanarak daha etkili optimizasyonlar yapılabilir (Eyduran, 2016).

Elde edilen sonuçlar, bağımlı ve bağımsız değişkenlerin dağılımına ait varsayım gerektirmeyen ve kolay yorumlanabilen MARS uygulamalarının, ırk standartlarını ortaya koymak ve burada incelenen canlı ağırlık özelliğini olumlu etkileyen morfolojik özellikleri belirlemek için önemli bir seçenek olabileceğini göstermiştir. Bu tez çalışmasının ileride benzer konularda çalışacak bilim adamlarına yararlı olabileceği ümit edilmektedir.

## KAYNAKLAR

- Abayomi-Alli, A.A., Omiddiora, O.E., Olabiyisi, O.S., Ojo, A.J., 2015. Adaptive Regression Splines Models for Predicting Facial Image Verification and Quality Assessment Scores. *Balkan Journal Of Electrical & Computer Engineering*. **3(1):17-26**.
- Acciani, C., Fucilli, V., Sardaro, R., 2011. Data mining in real estate appraisal: A model tree and multivariate adaptive regression spline approach. *Aestimium* **58:27-45**.
- Adamczyk, K., Zaborski, D., Grzesiak, W., Makulska, J., Jagusiak, W., 2016. Recognition of culling reasons in Polish dairy cows using data mining methods. *Computers and Electronics in Agriculture*. **127: 26-37**.
- Ankaralı, H., Temel, O.G., Yazıcı, C.A., 2010. Regresyon Modellerine Alternatif Bir Yaklaşım (MARS). *Türkiye Klinikleri J.Biostat .2 (2): 58-66*
- Bakir, G., Keskin, S. Mirtagioglu, H., 2010. Determination of the effective factors for 305 days milk yield by regression tree (RT) method. *J. Anim. Vet. Adv.*, **9:55-59**.
- Çetin, F. A., Mikail, N., 2016. Hayvancılıkta Veri Madenciliği Uygulamaları. *Türkiye Tarımsal Araştırmalar Dergisi*. **3:79-88**.
- Dogan, I., 2003. Investigation of the factors which are affecting the milk yield in Holstein by CHAID analysis. *Ankara Univ. Vet. Fac. J.*, **50:65-70**.
- Doğan, B.B., Toprak, S., 2012. Diyarbakır Ticaret Borsası'nın Faaliyetleri ve Borsanın Üye Profilleri İle İlgili Veri Eksiklerinin MARS Analizi ile Düzeltilmesi. *C.Ü. İktisadi ve İdari Bilimler Dergisi*. **13(2):151-174**.
- Eyduran, E., Karakus K, Keskin S. Cengiz F. 2008. Determination of factors influencing birth weight using regression tree (RT) method. *J. Appl. Anim. Res.*, **34: 109–112**.
- Eyduran, E., Yılmaz, I., Kaygisiz, A., Aktas, Z.M., 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.*, 23:956-963.
- Eyduran, E., Yilmaz, I., Tariq, M.M., Kaygisiz, A., 2013b. Estimation of 305-d milk yield using regression tree method in Brown Swiss cattle. *The Journal of Animal & Plant Sciences*, 23:731-735.
- Eyduran, E., 2016. The Possibility of using Data Mining Algorithms in Prediction of Live Body Weights of Small Ruminants. *Adv Cal Anal Sci.*, 1:18-21.
- Eyduran, E., Akkuş, O., Kara, K.M., Tırnık, C., Tariq, M.M., 2017. Use of Adaptive Regression Splines (Mars) in Predicting Body Weight from Body Measurements in Mengali Rams. *International Conference On Agriculture, Forest, Sciences and Technologies.Cappadocia/ Turkey.*, 415.
- Grzesiak, W., Lacroix, R., Wójcik, J. Błaszczyk, P., 2003. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Can. J. Anim. Sci.*, 83:307-311.
- Grzesiak, W., Błaszczyk, P. and Lacroix, R., 2006. Methods of predicting milk yield in dairy cows- Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Comput. Electron. Agric.*, 54:69-83.
- Grzesiak, W., Zaborski, D., Sablik, P., Zukiewicz, A., Dybus, A., Kowska, S.I., 2011. Detection of cows with insemination problems using selected classification models. *Animal Science Papers and Reports*. 29(4):293-302.
- Grzesiak, W., Zaborski, D. 2012. *Examples of the Use of Data Mining Methods in Animal Breeding*. DOI:10.5772/50893.
- Karadaş, K., Ertürk, E.Y., Eyduran, E., Gürsoy, K.A., Tariq, M.M., 2017. Predictive Performances of Chaid and Mars Data Mining Algorithms in the Establishment of Relationship between Live Body Weight and Several Morphological Measurements of Indigenous Mengali Sheep and Its Economic Importance. *International Conference On Agriculture, Forest, Sciences and Technologies.Cappadocia/ Turkey.*, 117.

- Khan, M.A., 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 Sciences.*, **24**:120-126.
- Mendeş, M., Akkartal, E., 2009. Regression tree analysis for predicting slaughter weight in broilers. *Italian Journal of Animal Science* **8**:615-624.
- 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 Sciences*, **22(1)**:20-24.
- Nisbet, R., Elder, J. Miner, G., 2009. *Handbook of statistical analysis and data mining applications. Canada.*
- Tunay, B.K., 2010. Bankacılık Krizleri ve Erken Uyarı Sistemleri: Türk Bankacılık Sektörü İçin Bir Model Önerisi. *BDDK Bankacılık ve Finansal Piyasalar* **4(1)**:9-46.
- Tunay, B.K., 2011. Türkiye'de Durgunlukların MARS Yöntemi ile Tahmini ve Kestirimi. *Marmara Üniversitesi İ.İ.B.F. Dergisi.* **3(1)**:71-91.
- Yakubu, A., 2012. Application of Regression Tree Methodology in Predicting the Body Weight of Uda Rams. *Animal Science and Biotechnologies*, **45 (2)**:484-490.
- Zakeri, F.I., Adolph, L.A., Puyau, R.M., Vohra, A.F., Butte, F.N., 2010. Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents. *J Appl. Physiol.*, **108**:128-136.
- Zhang, W., Goh, A.T.C., Zhang, Y., 2016. Multivariate Adaptive Regression Splines Application for Multivariate Geotechnical Problems with Big Data. *Geotech Geol. Eng.*, **34**:193-204

## ÖZGEÇMİŞ

1981 yılında Van'ın Özalp İlçesinin Gözdeğmez köyünde doğdu. İlköğretim ve Ortaöğretimini Iğdır ilinde tamamladı. 2001 yılında İnönü Üniversitesi Adıyaman Fen/Edebiyat Fakültesi Matematik bölümünde lisans öğrenimine başladı ve 2005 yılında mezun oldu. Aynı yıl Yüzüncü Yıl Üniversitesi Tezsiz Yüksek Lisans öğrenimine başladı. 2008 yılında mezun oldu. 2012 yılında Iğdır sosyal yardımlaşma ve dayanışma vakfını işçi olarak işe girdi. Halen aynı işte çalışmaktadır. Evli ve 3 çocuk babasıdır.

