



## Veteriner hekimliği alanında karar ağaçları uygulamalarının incelenmesi

Özgecan KORKMAZ AĞAOĞLU<sup>1,2a\*</sup>, İsmayil Safa GÜRCAN<sup>2,b</sup>

<sup>1</sup> Burdur Mehmet Akif Ersoy Üniversitesi, Veteriner Fakültesi, Genetik Anabilim Dalı, Burdur, Türkiye,

<sup>2</sup> Ankara Üniversitesi, Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Ankara, Türkiye

0000-0002-7414-1725<sup>a</sup>; 0000-0002-0738-1518<sup>b</sup>

### MAKALE BİLGİSİ /

### ARTICLE INFORMATION:

#### Geliş / Received:

12 Kasım 22

12 November 22

#### Revizyon/Revised:

12 Ocak 23

12 January 23

#### Kabul / Accepted:

09 Mart 23

09 March 23

#### Anahtar Sözcükler:

Veri madenciliği

Karar ağaçları

Veteriner hekimlik

#### Keywords:

Data mining

Decision tree

Veterinary medicine

### ÖZET

Bilimsel araştırmalar sonucunda elde edilen verilerin analiz edilmesinde istatistiksel yöntemler önemli birer araçtır. Bununla birlikte; elde edilen verinin çok büyük olması gibi durumlarda klasik istatistiksel yöntemler yetersiz kalabilmektedir. Teknolojinin hızla gelişmesi ve bilgilerin depolanabilme kapasitelerinin artması, bilginin önemini daha da arttırmıştır. Bilginin önemli hale gelmesi, toplanan verinin büyük olması ve klasik istatistiksel yöntemlerin bu veriyi analiz etmede yetersiz kalması ise veri madenciliği gibi yöntemlerin doğmasına neden olmuştur. Veri madenciliği, dijital platformlarda depolanan devasa büyüklükteki veriler arasındaki örüntülerin değerlendirilmesi, çıkarımlar yapılması ve bunun sonucunda da anlamlı bilgiler elde edilmesi için uygulanan analizler olarak tanımlanmaktadır. Veteriner hekimliği; hayvan yetiştiriciliği, gıda güvenliği, gıda kalitesinin belirlenmesi, hayvan hastalıklarının yayılımı, hastalıkların teşhis ve tedavisi gibi birçok konuda veri üretimi nedeniyle veri madenciliğinin uygulanabileceği bir alandır. Bu derlemede veteriner hekimliği alanında son yıllarda yaygın bir şekilde kullanılmaya başlanan ve önemli bir sınıflandırma modeli olan karar ağaçları modelleme yönteminin içeriği ve kullanım alanlarının tanıtılması amaçlanmıştır.

### *Examination of decision trees applications in the veterinary medicine*

### ABSTRACT

Statistical methods are important tools in the analysis of data obtained as a result of scientific research. However, in cases where the data obtained is very large, classical statistical methods may be insufficient. The rapid development of technology and the increase in the storage capacity of information have increased the importance of information even more. The fact that information has become important, the data collected is large, and classical statistical methods are insufficient to analyze this data has led to the emergence of methods such as data mining. Data mining is defined as the analysis applied to evaluate the patterns among the huge data stored on digital platforms and to make inferences to obtain meaningful information. Veterinary science is an area where data mining can be applied because it produces data on many subjects such as animal husbandry, food safety, determination of food quality, the spread of animal diseases, diagnosis and treatment of diseases. This review, it is aimed to introduce the content and usage areas of the decision tree modeling method, which has been widely used in the field of veterinary medicine in recent years and is an important classification model.

©2023 The Authors.

Published by Veteriner

Hekimler Derneği. This is

an open access article

under CC-BY-NC license.

(<https://creativecommons.org/licenses/by-nc/4.0>)



## 1. Giriş

Bilgi sistemleri ve teknolojinin artan hızla gelişimi ile saklanan ve işlenebilen veri miktarının artması, verilerin farklı analiz yöntemleri ile yorumlanması ihtiyacını gündeme getirmiş, dünya literatüründe “Data Mining” (1) olarak ifade edilen “Veri Madenciliği” kavramını ortaya çıkarmıştır.

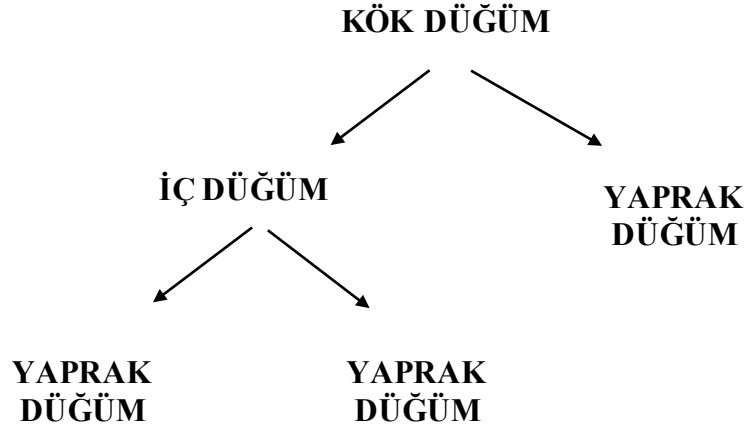
Veri madenciliği, dijital platformlarda depolanan devasa büyüklükteki veriler arasındaki örüntüleri keşfederek çıkarımlar yapmak, farklı açılardan analiz ederek anlamlı bilgiler elde edilmesi için uygulanan analizler olarak tanımlanmakla birlikte, son yıllarda strateji odaklı karar desteği sağlamaya yönelik kullanılmaktadır (2-6). Bu bağlamda sağlık bilimleri alanında üretilen ve farklı birçok disiplinden toplanan büyük veri de oldukça dinamik ve karmaşık bir yapıya sahip olması nedeniyle veri madenciliği için kullanılabilir bir alan oluşturmaktadır (7). Veteriner hekimliği alanı da birçok farklı disiplini bünyesinde barındırması ve hayvan yetiştiriciliği, gıda güvenliği, hayvan hastalıklarının yayılımı, teşhis ve tedavisi gibi birçok konuyu kapsamaması nedeniyle veri madenciliğinin uygulanabileceği bir alandır. Veri madenciliği; veteriner hekimlere ve araştırmacılara, araştırılan konu ile ilgili yeni bilgi üretme, hastalık teşhis ve tedavi süreçlerinde daha isabetli kararlar verme, herhangi bir araştırma konusunda yapılacak tahmini modellemelerin daha gerçekçi olmasını sağlama ve hayvan ıslahı programları gibi programların planlanmasında yarar sağlamaktadır (8-10). Bunun yanı sıra veri madenciliği veteriner pratikte; kanatlı mortalitelerinin tahmini, ıslah programlarının geliştirilmesinde yardımcı veriler oluşturulması, mastitis gibi ekonomik önemi olan hastalıkların belirlenmesi, çeşitli hastalıklar için teşhis modelleri ve tedavi stratejilerinin geliştirilmesi, genomik veriler kullanarak hayvanların kimlik teyidi çalışmalarının yapılması, farklı parametreler kullanılarak et kalitesinin değerlendirilmesi, miRNA’ların hedef genlerinin tahmin edilmesi, hayvan davranışları, tür tayini, ekonomik önemi olan verim özelliklerinin tahmin edilmesi gibi birçok farklı konuda kullanım alanı bulmaktadır. Bu amaçlarla kullanılan veri madenciliği için geliştirilmiş yöntemler genel olarak; sınıflandırma (Classification), kümeleme (Clustering) ve birliktelik kuralları (Association Rules) şeklinde gruplandırılmaktadır (3). Bu yöntemler içinde en sık kullanılan yöntem ise sınıflandırma yöntemidir. Başlıca sınıflandırma modelleri; Karar Ağaçları (Decision Trees) (11), Yapay Sinir Ağları (Artificial Neural Networks) (12), Genetik Algoritmalar (Genetic Algorithms) (13), Naive-Bayes (14) ve Lojistik Regresyon (Logistic Regression)’dur (15).

Bu derlemede veteriner hekimlik alanında son yıllarda yaygın bir şekilde kullanılmaya başlanan ve önemli bir sınıflandırma modeli olan karar ağaçları modelleme yönteminin içeriği ve kullanım alanlarının tanıtılması amaçlanmıştır.

## 2. Karar Ağaçları

Veri madenciliğinde güvenilir ve sıklıkla tercih edilen sınıflandırma yöntemlerinden biri olan karar ağaçları, ilk defa Breiman ve arkadaşları (11) tarafından önerilmiştir. Bu yöntem kullanılarak verilerin sınıflandırılması iki aşamalı olarak gerçekleştirilmektedir. Birinci aşama öğrenme aşaması, ikinci aşama ise test verisi aşaması olarak isimlendirilmektedir. Öğrenme aşamasında; bilinen bir öğrenme veri seti, model oluşturmak amacıyla sınıflandırma algoritması ile belirlenmektedir. Öğrenilen model, sınıflandırma kurallarını oluşturur ve karar ağacı şeklinde ifade edilmektedir. Test verisi aşaması ise sınıflandırma kurallarının doğruluğunu belirlemek amacıyla kullanılmaktadır. Sınıflandırma kurallarının doğruluğu kabul edilebilir düzeyde ise elde edilen kurallar elde edilen yeni verilerin sınıflandırılması için kullanılabilir (16,17). Karar ağaçları aynı zamanda regresyon amacıyla da kullanılmaktadır.

Karar ağaçlarında her bir değişken bir düğüm (node) ile temsil edilmektedir. Karar ağacında en üst kısım kök, en alt sınıflandırmayı sağlayan düğüm ise yaprak (leaf) düğüm olarak isimlendirilir. Kök ile yaprak düğümler arasında kalan kısımlara ise dal (branch) adı verilir (18). Karar ağaçlarının genel yapısı Şekil 1’deki gibi görselleştirilebilir.



Şekil 1: Karar ağacı temel yapısı (Adua ve ark, 2021 (19)'den Türkçe'ye uyarlanmıştır)

*Figure 1: The basic structure of decision tree (adapted to Turkish from Adua et al., 2021 (19))*

Karar ağaçları oluşturulması için farklı algoritmalarından yararlanılabilmektedir. Seçilen algoritmaya göre oluşturulan karar ağacının şekli de değişebilmektedir (20). Karar ağaçlarında kökün ne olacağı, dallanmaların hangi kriterlere göre yapılacağı kullanılan algoritmaya göre belirlenmektedir. Karar ağacı oluşturmak için kullanılan ilk algoritma Morgan ve Sonquist, 1963 (21) tarafından bildirilen otomatik etkileşim belirleme analizi (Automatic Interaction Detector; AID) algoritmasıdır. Daha sonra bu alanda yapılan çalışmalar sonucunda; otomatik Ki-kare etkileşim belirleme analizi (Chi-squared automatic interaction detector; CHAID) (22), sınıflama ve regresyon ağaçları (classification and regression trees; CART) (11), tekrarlı ikililikçi ağacı (iterative dichotomiser 3; ID3 ağacı) (23), ayrıntılı otomatik Ki-kare etkileşim belirleme analizi (Exhaustive Chi-squared automatic interaction detector; CHAID) (24), çok değişkenli adaptif regresyon eğrileri (multivariate adaptive regression splines; MARS) (25), C4.5 karar ağacı (26), C5 karar ağacı (26), CAL5 karar ağacı algoritması (27), QUEST'te denetimli öğrenme (supervised learning in QUEST, SLIQ) (28), karar ağaçlarının ölçeklendirilebilir paralel indüksiyonu (scalable parallelizable induction of decision trees; SPRINT) (28), çabuk, önyargısız, etkin istatistiksel ağaç, (quick, unbiased, efficient statistical tree; QUEST) (29), rastgele orman (Random Forest) (30), arttırılmış ağaçlar (boosted trees) (31), J48 ağacı (C4.5'in Weka'daki uygulaması) (32), döndürme ağacı (rotation forest) (33) gibi farklı algoritmalar geliştirilmiştir.

Karar ağaçlarının oluşturulması aşamasında üzerinde durulması gereken en kritik aşama; veri setinde bulunan her bir değişkenin sınıflandırılmasını sağlayacak dallanmanın hangi kritere göre yapılacağını belirlenmesidir (20). Ağaç oluşturulurken belirsizliği en fazla olan değişkene karar verilmesi ve ağacın kök düğümü için test edilmesi önem arz etmektedir. Değişkenin belirlenmesine yönelik olarak ilgili literatürde entropiye dayalı olan; bilgi kazancı ve bilgi kazanç oranı (34-36), gini kriteri (11), twoing kuralı (11) ve ki-kare olasılık (37) tablo istatistiği gibi farklı metodolojilerin kullanılabileceği bildirilmiştir.

Ağaç oluşturulması sırasında bazen basit bir veri kümesinden büyük bir ağaç oluşabilmektedir. Karar ağacı oluştururken istenmeyen bir durum olarak oluşan bu olaya aşırı öğrenme (**overfitting**) adı verilmektedir (38). Aşırı öğrenme; veri yığını üzerinde gürültü olması ve/veya seçilmiş veri kümesinin o durumu temsil etmemesi sonucunda oluşmaktadır. Ayrıca; aşırı öğrenme test veri setine ait hata oranının yüksek çıkmasına neden olabilmektedir. Ağaç oluşturma amacıyla kullanılan algoritmalar sıklıkla aşırı öğrenme oluşturabilmektedirler. Oluşan ağacın çok büyük olması bu etkiyi fazlasıyla arttırmaktadır. İstenmeyen bir durum olarak şekillenen aşırı öğrenmenin önüne geçmek için çeşitli yöntemler uygulanmaktadır. Bu yöntemlerden biri budama (pruning) yapmaktır. Budama işlemi iki farklı şekilde gerçekleştirilebilmektedir (34). Bu yöntemlerin ilki, ağaç yapısı oluşturulması sırasında belli bir oranda büyüyen ağacın daha çok büyümesine engel olmak için bölünmeyi durdurup ön budama (prepruning) işlemi yapmaktır. Bu yöntemde dezavantaj olarak çelişkili sonuçlar elde edildiğini bildiren görüşler bulunmaktadır (11). Budama için kullanılan diğer

yöntem ise ağaç oluşturulduktan sonra şekillendirilmiş olan bölünme noktalarının çıkarılması ile yapılan son budama işlemidir (postpruning).

Bahse konu edilen veri madenciliği yöntemlerinin uygulanabilmesi için kullanılması gereken verilerin dijital ortamda olması gerekmektedir. Birçok farklı alanda yaşanan dijitalleşmeye veteriner hekimlik alanı da ayak uydurmuş ve hızlı bir dijitalleşme sürecine girilmiştir. Örneğin; veteriner klinikleri, hayvan hastaneleri, süt ineği işletmeleri, kanatlı hayvan işletmeleri ve yem üreticileri gibi birçok farklı alanda artık veriler periyodik olarak dijital ortamlara aktarılmakta ve yazılım programları ile işlenmektedir. Bu gelişmeler sonucunda veteriner hekimlik alanında da veri madenciliği çalışmalarının ham maddesini oluşturan dijital veri kaynağı oluşmuş ve oluşmaya devam etmektedir. Verilerin toplanmasında yaşanan bu hızlı değişim ile eş zamanlı olarak da veteriner hekimlik alanında veri madenciliği üzerinde yapılan çalışmalar ivme kazanmıştır. Bu bağlamda; veteriner hekimliği alanında kullanılan veri madenciliği yöntemlerinden biri olan karar ağaçlarının kullanıldığı çalışmalarından bir kısmı bir sonraki bölümde özetlenmiştir.

### 3. Veteriner Hekimliği Alanında Karar Ağacı Uygulamaları

Vale ve ark. (39), ısı dalgası insidansının broiler tavuklarının mortaliteleri üzerindeki etkisini hesaplamak için karar ağaçlarını (J48 algoritması) kullanmışlardır. Bu amaçla broiler mortalitesi ile iç ve dış çevresel faktörleri (sıcaklıklar, nispi nem, nemlilik indeksi vb gibi) kullanarak bir model oluşturup; 29 ve 42 günlük broilerler için zararlı çevresel koşul aralıklarını belirlemişlerdir. Bu verileri kullanarak da broiler mortalitelerinin karar ağaçları ile tahmin edilebileceğini ortaya koymuşlardır.

Piwczyński, (40) 2-8 yaşlarında 6586 Polonya Merinos koyununda anaç koyunların büyütmüş olduğu yavruların sayısında değişime neden olan faktörleri belirlemek için sınıflandırma ağaçlarını (Gini indeksi) kullanmıştır. Yavru sayısındaki varyasyon; koyun canlı ağırlığı, üretim sektörü, doğum ve yetiştirme tipi, doğum ve kuzulama yılı, koyun yaşı ve sürü odaklı olarak değerlendirilmiştir. Çalışmada sürü faktörünün yavru sayısını etkileyen oldukça önemli bir etken olduğu ayrıca veri alt kümelerini oluşturan en önemli faktörün de canlı ağırlık olduğu belirlenmiştir.

Pinzón-Sánchez ve ark. (41) yaptıkları çalışmada; laktasyonun erken döneminde farklı patojenlerin etkisiyle oluşan hafif veya orta şiddette klinik mastitis olgularında uygulanan meme içi tedavi sürelerinin ekonomik etkilerinin değerlendirilmesi için karar ağacı geliştirmişlerdir. Olgular meme lobu düzeyinde değerlendirilmiştir. Araştırma sonucunda, karar ağacının ticari sütçü sürülerde ekonomik olarak en uygun tedavinin seçilmesinde kullanılabilecek etkili bir yöntem olduğu sonucuna varılmıştır.

Židek ve ark., (42) Avusturya ve Slovakya orijinli 412 baş Pinzgau sığırının DNA'ları ile hayvanlara ait kimlik kontrolü/teyidi amacı ile sınıflandırma modeli geliştirmişlerdir. Çalışmada toplam 412 baş sığır orijinlerine göre 1. sınıf (n=346) ve 2. sınıf (n=66) olmak üzere ayırmışlardır. Hayvanların genetik örüntüleri 20 farklı veri madenciliği metodu ile modellenmiştir. Üç model yüksek belirleme başarısı değerlerine göre seçilmiştir. En küçük algoritma hatası karar ağaçları uygulanan modelde bulunmuştur. Sonuç olarak; veri madenciliği yöntemlerinin kullanımı ile genetik örüntülere dayalı bilinmeyen örneklerin sınıflandırılabilceği gösterilmiştir.

Wylie ve ark., (43), atlarda laminitis ile diğer topallık nedenlerinin prevalanslarını karşılaştırmak ve laminitisleri ayırt edebilmek için karar ağaçlarının kullanılabilirliklerini çalışmışlardır. Çalışmada önce atların göstermiş olduğu semptomlar değerlendirilerek, atlar laminitis olan ve olmayan olarak ayrılmıştır. Ayrıca; veriler oluşturulan laminitis değerlendirme formuna işlenmiş, duruş özellikleri, topallık durumu, etkilenen ayaklar ve laminitise bağlı oluşan akut ya da kronik belirtiler olmak üzere 5 bölüme ayrılmıştır. Daha sonra semptomlara bakarak konan teşhisler ve form değerlendirmeleri karşılaştırılmış, her klinik semptom ile olgu arasındaki ilişki %95 güven aralığında düzeltilmiş olasılık oranları kullanılarak lojistik regresyon modeli ile değerlendirilmiştir. Kaydedilen klinik semptomlar kullanılarak i) topallık, ii) duruş, iii) etkilenen ayaklar, iv) akut semptomlar, v) akut ve kronik semptomlar olmak üzere 5 adet ön ağaç modeli oluşturulmuştur. Daha sonra beş ön ağaç değerlendirmesinden laminitis için en büyük belirteç olarak tanımlanan değişkenler kullanılarak iki birleşik ağaç modeli oluşturulmuştur. İlk kombine ağaç modeli; kronik laminitis belirtisi olmayan atlarda akut klinik belirtiler gözlemlenen ayaklar ve duruş özelliklerinden oluşturulmuş, ikinci kombine ağaç modeli ise kronik laminitis olan atlarda akut ve kronik klinik belirtiler gözlemlenen ayaklar ve duruş özelliklerinden oluşturulmuştur. Sonuç olarak; karar ağacı modellerinin laminitis gibi bazı

hastalıkların tespitinde kullanılabileceği ancak geliştirilmesi gerektiği belirlenmiştir.

Velásquez ve ark. (44) yaptıkları çalışmada; karar ağacı yöntemini kullanarak sığır etinin hiperspektral görüntü analizine dayalı olarak mermerleşme derecesini belirlemeyi amaçlamışlardır. Bu amaçla; 29 dişi ve 6 erkek sığır karkasında *M. longissimus dorsi* kasından örnek alınmıştır. Alınan örnekler -18 °C'de 24 saat süreyle tutulmuş ve 2 cm kalınlığında parçalar kesilmiştir. Kesitler 2 °C'de 24 saat bekletilerek çözdürülmüş ve hiperspektral görüntüleme kamerası (Pica XC, Resonon Inc., ABD) ile taranmışlardır. Örnekler 0.5 cm/s hızla satır satır taranarak hiperspektral görüntüler elde edilmiştir. Görüntüler; referans değerler kullanılarak MATLAB 7.1 R2010a yazılımı ile düzeltilmiştir. Alınan kesitlerde etin farklı bölümleri (yağlı ve yağsız alanlar) MATLAB 7.1 R2010a programı kullanılarak tanımlanmıştır. Elde edilen sığır eti görüntülerinde mermerleşme dağılımını analiz edebilmek için görüntülerde oluşturma (building) ve sınıflandırma işlemleri yapılmıştır. Oluşturma aşamasında kullanılmak üzere alınan hiperspektral görüntülerden rastgele 10 görüntü seçilmiştir. Bu görüntüler üzerinde yağ ve et pikselleri seçilmiştir. Seçilen piksellerin spektral profilleri ile bir veri tabanı oluşturulmuştur. Veri setinden aykırı değerler ayıklanmıştır (45). Sonuç olarak karar ağacında kullanmak için 20.000 piksellik (10.000 yağ, 10.000 et) bir veri seti oluşturulmuştur. Karar ağacı; alınan görüntülerdeki piksellerin spektral verilerine dayanarak etin yağlı mı yağsız mı olduğunu belirlemek amacıyla oluşturulmuştur. Son olarak tüm örnekler; geleneksel değerlendirme yöntemi (46) ile 25 farklı kişi tarafından değerlendirilmiştir. Hiperspektral görüntülerin değerlendirilmesi ve geleneksel değerlendirme yöntemi ile derecelendirilen örnekler karşılaştırıldığında elde edilen verilerin istatistiksel bir fark oluşturmadığı görülmüştür. Bu nedenle; et mermerleşmesinin derecelendirilmesinde karar ağacına dayalı değerlendirme sisteminin; hızlı, makine öğrenmesi yolu ile geliştirilebilir ve yüksek doğruluk oranına sahip olduğu sonucuna varılmıştır.

Gagaoua ve ark. (47), karar ağacı temelli öğrenme yöntemi kullanarak karkas özelliklerinin tahmin edilebilirliğini araştırmışlardır. Çalışmada farklı ırklardan (Charolais n=116, Salers n=93, Limousin n=74, Blond d'Aquitaine n=25) 308 sığır kullanılmıştır. Kullanılan sığırlar; 10 farklı yetiştirme faktörü kullanılarak karakterize edilmiştir. Bu faktörler sırasıyla; kesim yaşı (ay), besi süresi (gün), besi başlangıcındaki canlı ağırlık (kg), besi sonu canlı ağırlık (kg), günlük kuru madde alımı (KM; kg/gün), kaba yem ve konsantre yem oranları (rasyondaki KM oranları), günlük enerji alımı (Mcal/gün), besi dönemindeki ortalama günlük canlı ağırlık artışı, besleme verimliliği (günlük canlı ağırlık artışı/günlük kuru madde alımı) olarak belirlenmiştir. Çalışmada kesim sonrası 2-4 °C'de 24 saat bekletilen karkasların *M. longissimus thoracis* kası çıkartılmıştır. Kasın çevresini saran bağ doku diseke edilmiş ve kalan kas 2 cm kalınlığında dilimlenmiştir. Kesilen her bir dilim plastik vakumlu pakete alınıp 14 gün 4°C'de olgunlaştırılmıştır. Olgunlaştırma sonrasında -20°C'de dondurulmuştur. Karkaslar AB karkas sınıflandırma kriterlerine göre sınıflandırılmışlardır (EU Commission Regulation EC 1249/2008, EUROP). Bu kriterlere göre karkaslar E (çok kaslı) harfinden P harfine doğru (az kaslı) 5 farklı tipte derecelendirilmişlerdir. Ayrıca yağlılık için de 1'den (yağsız) 5'e (çok yağlı) kadar bir numara verilmiştir. Karkas kompozisyonunun hesaplanabilmesi için; kas/karkas oranı, yağ/karkas oranı ve kemik/karkas oranı belirlenmiştir. Etin gevrekliğinin belirlenebilmesi için; 14 gün olgunlaştırılıp -20°C'de saklanan etler pişirilmeden 48 saat önce 5°C'lik soğutuculara alınıp çözdürülmüştür. Etler 300°C'ye ısıtılmış ızgara üzerine konmuş ve 2 dakika süreyle merkezi ısıları 55°C'ye çıkana kadar pişirilmiştir. Tadım ve skorlama için pişirilen etler 12 değerlendiriciye sunulmuştur. Skorlama 0 (aşırı sert) ile 10 (aşırı yumuşak) arasında puan verilerek yapılmıştır. Alınan veriler SAS v. 9.4 (SAS Institute Inc. Cary, NY, ABD) ve XLSTAT 2017.19.6 (AddinSoft, Paris, Fransa) yazılım programları kullanılarak analiz edilmiştir. Öğrenme araçları olarak; hiyerarşik küme analizleri (hierarchical cluster analysis, HCA), k-ortalamarı (k-means) ve k-medoids kümeleme algoritmaları kullanılmıştır. Ölçülen gevreklik puanlarının; Z puanları hesaplanarak herhangi bir kümeleme analizi yapılmadan önce ortalamarı alınmıştır. Bu ortalamalar; ırk veya deneyden bağımsız olarak ortalamanın karşılık geldiği hayvanda her bir özellik için yapılan gözlemin sapmasını temsil etmektedir. K-ortalamarı, bir veri kümesini otomatik olarak k kümelerine bölmek için kullanılan yinelemeli bir kümeleme yöntemidir. Bu algoritma iki ayrı aşamadan oluşmaktadır. İlk aşama; her küme için bir tane k merkez noktası tanımlanmasıdır. İkinci aşama; veri kümesine ait her noktayı alıp en yakın merkezi değer ile ilişkilendirmektir. Hiyerarşik küme analizi ise; veriler arasında bir hiyerarşi oluşturularak veri kümesinden ikili bir ağaç oluşturmayı sağlamaktadır. Hiyerarşik küme analizi, veriler arasındaki benzerliği veya farklılığı ölçmeye dayalı bir şekilde grupları homojen ya da harici olarak izole edilmiş olarak sınıflandırmaya olanak sağlamaktadır. Yapılan kümelemelerin sonuçlarının doğrulanması için siluet genişliği (S<sub>i</sub>) kriteri kullanılmıştır. S<sub>i</sub> değeri -1 ile +1 arasında

değişen bir değerdir.  $S_i$  değerinin 1'e yaklaşması gözlemlerin iyi kümelendiğini, 0'a yaklaşması gözlemlerin iki küme arasında bulunduğunu, -1'e yaklaşması ise gözlemlerin muhtemelen yanlış kümelere bölündüğünü göstermektedir. Daha sonra sınıflandırma analizlerinin yapılabilmesi için; besi dönemi özellikleri, karkas özellikleri veya her ikisinin de kullanılmasıyla, CHAID, CART ve QUEST yöntemlerinden yararlanılarak farklı karar ağaçları oluşturulmuş ve k-ortalamları tahmin edilmeye çalışılmıştır. Çalışmada en iyi sonuçlar CART yönteminden elde edilmiştir. Çalışma sonucunda farklı yetiştirme faktörlerinin ve karkas özelliklerinin karar ağaçlarında güvenilir ve uygulanabilir olduğu belirlenmiştir.

Ekiz ve ark., (48) yaptıkları çalışmada; oğlak karkaslarında doku kompozisyonunun değerlendirilmesinde, klasik aşamalı regresyon (step-wise regression, SWR) analizi ile CHAID algoritması (karar ağacı) ve yapay sinir ağı (artificial neural network, ANN) yöntemlerini karşılaştırmışlardır. Bu amaçla; 57 baş oğlak kullanılmıştır. Tahmin değişkeni olarak kesim öncesi ağırlığı, sıcak karkas ağırlığı, karaciğer ağırlığı, omental yağ ağırlığı, boş vücut ağırlığı, soğuk karkas ağırlığı, karkas uzunluğu, bacak uzunluğu, göğüs genişliği, göğüs derinliği, göğüs çevresi, but genişliği, but çevresi, internal karkas uzunluğu, arka bacak uzunluğu ölçülmüştür. Ayrıca; sıcak karkas ağırlığının internal karkas uzunluğuna oranı hesaplanarak karkas kompaktlığı, bacak uzunluğunun arka bacak uzunluğuna oranı belirlenerek arka bacak kompaktlığı, göğüs genişliğinin göğüs derinliğine oranı hesaplanarak göğüs kafesi yuvarlaklık indeksi belirlenmiştir. Ayrıca son torasik vertebra ile ilk lomber vertebra arasında yer alan longissimus dorsi kasının kesit alanı ölçülmüştür. Yapay sinir ağı, karar ağacı ve klasik çoklu regresyon analizleri ile karkas dokusu kompozisyonunun tahmini için analizler yapılmıştır. Karar ağacı algoritması SPSS Modeler v.18. kullanılarak tasarlanmıştır. İlk analizde; kas, kemik, deri altı yağ, kaslar arası yağ veya toplam yağ oranlarını tahmin eden en iyi karar ağacı modelini belirlemek için CHAID ve CART algoritmaları karşılaştırılmıştır. Karar ağacında kullanmak üzere CHAID algoritmasının kullanılmasının uygun olacağı belirlenmiştir. Çalışma sonunda kas oranını en iyi tahmin eden sistemin yapay sinir ağı olduğu ortaya konmuştur. Kemik oranı açısından karar ağacı ve yapay sinir ağının benzer sonuçlar verdiği tespit edilmiş klasik regresyon modelinin düşük  $R^2$  değerlerine sahip olduğu bulunmuştur.

Tamura ve ark. (49), sığırların davranışları ile ivme ölçer sensörlerden elde edilen veriler arasındaki ilişkiyi incelemişlerdir. Bu amaçla; 38 baş sığıra üç eksenli ivme ölçerler takılmıştır. İvme ölçerden alınan sinyal verilerinden; yem yeme, geviş getirme ve yatma davranışları için ayrı ayrı veri setleri oluşturulmuştur. Tespit edilen her bir değer için; ileri geri, yukarı aşağı ve sağa sola hareketlerin birleşimi olan üç boyutlu hareketler için skalar değer hesaplanmıştır. Eşik değerlerin hesaplanabilmesi için yem yeme, geviş getirme ve yatma aktiviteleri seviyesi ile oluşturulan veri seti kullanılarak karar ağacı öğrenmesi yapılmıştır. Karar ağacı oluşturulurken CART algoritması (11) kullanılmıştır. Çiftliklerden alınan veri setleri, oluşturulan karar ağacına uygulandığında tüm verilerin doğru sınıflandırıldığı, duyarlılık ve özgüllüğün %100 olduğu belirlenmiştir. Çalışma sonucunda ineklerin boynuna takılan üç eksenli ivme ölçer ile yem yeme, geviş getirme ve yatma davranışlarının sınıflandırılması, seviyesi ve varyasyonlarının makine öğrenmesi ve karar ağaçları ile belirlenebileceği ortaya konulmuştur. Bu yöntemin ileride süt ineği yetiştiriciliğinde sürü idaresi için önemli ve faydalı bir araç olduğu ortaya konulmuştur.

Piwczyński ve ark. (50), otomatik sağım sistemi kullanılan işletmelerde karar ağaçları kullanarak süt verimlerinin tahmin edilebilirliğini araştırmışlardır. Bu amaçla; sağım robotu kullanılan 27 süt ineği işletmesinden toplam 3778 baş inekten veri alınmıştır. 2011 ile 2014 yılları arasında doğum yapan ineklerin ardışık süt verimleri analiz edilmiştir. Süt verimini etkileyen faktörleri; laktasyon sayısı, buzağılama yaşı, sağıldığı ay, doğurduğu yıl, sağım durumu, buzağılama ve sağım mevsimi, mevsim durumu, laktasyon ayı, barınak tipi, altlık materyali, sağım sıklığı, sağım hızı, sağılan meme lobu sayısı, sağım robotu başına düşen inek yoğunluğu olarak belirlenmiştir. Belirlenen bu faktörlerin süt verimi üzerindeki olası etkileri karar ağaçları kullanılarak analiz edilmiştir. Kullanılan 36.005 verinin %60'ı öğrenme veri seti, %40'ı ise test veri seti olarak kullanılmıştır. Ağacın son düğümü/dalının boyutunun 30 gözlemden az olmaması, derinliğinin (dalların sayısı) ise 5'ten derin olmayacağı varsayılmıştır. Bu varsayımla karar ağacı; minimum sayıda örnek (30 inek) veya maksimum sayıda bölüm (5 inek) elde edilene kadar bölünmüştür. Test veri seti içindeki rastgele ilişkileri yansıtan verilerin hatalı olarak öğrenme veri setine aşırı uyum göstermesini önleyebilmek için yaprak boyutu ve derinliği için kriterler belirlenmiştir. Test edilen değişken olan süt veriminin süreklilik arz eden bir özellikte olması nedeniyle F testi istatistiği ve varyans azaltma olmak üzere iki farklı bölme kriteri kullanılmıştır. Belirtilen bu bölme kriterlerine göre karar ağacı modellerinin ortalama kare hatasına dayalı olarak

yüksek süt verimini tahmin etme kalitesi araştırılmıştır. Çalışma sonunda karar ağaçları robotik sağım sistemine sahip olan işletmelerde süt verimini etkileyen en önemli faktörleri önem sırasına göre; sağım sıklığı, bir sonraki laktasyon, sağımın yapıldığı ay ve altık/yatak tipi olarak belirlemiştir. Ayrıca yine karar ağacı yöntemi ile bahsedilen faktörler arasında birçok interaksiyon olduğu ortaya konulmuştur. Karar ağaçları yönteminin; yetiştiricinin sahip olduğu sürünün maksimum süt üretimini sağlaması, çevresel faktör ve parametrelerin belirlenmesi ve kontrol altında tutulmasına olanak vereceği belirtilmiştir. Elde edilen veriler sağım süresinin sağım üzerinde önemli bir etkisi olduğunu göstermektedir. Bu nedenle sağım süresinin seçim kriterlerine alınması gerektiği belirtilmiştir. Karar ağaçları; barınak koşullarının önemini de ortaya koymuştur. Elde edilen sonuçlara göre; sığ durakların süt verimini azalttığı bildirilmiştir. Derin yataklı durakların verimler yönünden yararlı olduğu belirlenmiştir. Bu veriyi doğrulamak ve çevresel faktörleri elimine etmek için çevresel faktörlerden bağımsız olarak çalışmanın farklı sürülerde yapılması gerektiği bildirilmiştir.

Pascottini ve ark., (51), yaptıkları çalışmada ineklerde karar ağaçlarından yararlanarak geçiş dönemi koşulları ve hastalıkların fertilitite ile ilişkilerini değerlendirmeyi amaçlamışlardır. Makine öğrenmesi yöntemleri; mastitis kontrol programları (52), östrus belirleme (53, 54), üreme takibi (55, 56) vb. gibi değerlendirmelerde süt ineği sürü yönetim programlarında kullanım alanı bulmaktadırlar. Makine öğrenmesi veya karar ağaçlarının kullanılmasının; süt ineği yetiştiriciliğinde büyük önem taşıyan geçiş dönemi hastalıkları ile çok sayıda farklı faktörün interaksiyonunun belirlenmesinde yardımcı olabileceği düşünülmektedir. Bu bağlamda yapılan çalışmada; geçiş dönemi hastalıkları ile karmaşık olmayan, karmaşık ve toplam geçiş dönemi hastalıklarının kısa dönem (sağılan gün sayısı 120'ye kadar) ve uzun dönem (sağılan gün sayısı 210'a kadar) döl tutmama riski ile ilişkisinin araştırılması amaçlanmıştır. Bu amaçla; konvansiyonel Kaplan-Meier ve Cox orantısal modelleme ile karar ağacı ve makine öğrenmesi yöntemleri kullanılmıştır. Çalışmada toplam 2450 ineğe ait veri işlenmiştir. Karar ağacı; tekil, karmaşık ve toplam geçiş dönemi hastalıkları, buzağılama mevsimi, parite (primipar veya multipar) ve ilk 60 günlük süt verimi (düşük veya yüksek) verileri kullanılarak oluşturulmuştur. Çalışma sonunda konvansiyonel yöntem ve karar ağacı yöntemi karşılaştırıldığında sonuçların benzer olduğu belirlenmiştir. Ancak makine öğrenmesinin özellikle çok sayıda değişkenin kullanıldığı ya da eksik değerlerin bulunduğu konvansiyonel yöntemin uygulanmasının çok mümkün olmadığı durumlarda daha geçerli bir yaklaşım olduğu sonucuna varılmıştır.

Romero ve ark. (57) yaptıkları çalışmada; sığır tüberkülozis hastalığının kontrolünde karar ağaçlarının kullanımını araştırmışlardır. Sığır sürülerinde tüberkülozis için risk faktörlerinin belirlenmesi hastalığın kontrolü için büyük önem taşımaktadır. Bu risk faktörlerinin ortaya konulmasında makine öğrenmesi ve karar ağaçları yöntemlerinin kullanılması ile hastalığın etiolojisinde yer alan ve doğrusal olmayan birçok ilişki ve etkileşim ortaya çıkartılabilir. Çalışmada 2016 yılında İngiltere'de bulunan sığır işletmelerinden alınan; sürülerin demografik özellikleri, kayıt sisteminde bulunan tüberkülozis geçmişleri, hayvan hareketleri, iklim değişiklikleri, yaban hayvanı varlığı, arazi sınıflandırma verileri kullanılmıştır. Veriler ArcMap (ESRI) yazılım programı kullanılarak sürü düzeyinde düzenlenmiştir. Karar ağacı algoritmasının hızını ve performansını iyileştirmek için önemli olmayan değişkenler çıkartılmıştır. Bu amaçla; her bir tahmin edici veri (predictor) ile sonuç arasında tek değişkenli lojistik regresyon analizi yapılmış ve  $p > 0.1$  olan değişkenler çıkartılmıştır (58). 0.79'un üzerinde bir korelasyon katsayısı ile yüksek korelasyona sahip değişkenler belirlenmiştir (59). Yüksek korelasyon katsayısına sahip çiftler arasından, bu tahmin edici veri ile diğer tüm değişkenler arasında en düşük ortalama korelasyona sahip olan predictor seçilmiştir (60). Yüksek oranda korelasyona sahip olan ve olmayan seçilmiş değişkenlerden veri setleri oluşturulmuştur. Sıfıra yakın varyasyon değerine sahip predictorler çıkartılmıştır. Karar ağacı modeli CART algoritması kullanılarak geliştirilmiştir (61). Çalışmada 52.668 sığır sürüsünden gelen veri işlenmiştir. Toplam 65 predictor kullanılmıştır. 2016 yılında İngiltere'de sığırlarda görülen tüberkülozis hastalığının insidensini belirlemek için 13 düğüm içeren bir karar ağacı oluşturulmuştur. Yüksek riskli alanların belirlenmesi için ise 19 düğümden oluşan bir karar ağacı oluşturulmuştur. Çalışma sonunda tüberkülozis ile ilişkili risk faktörlerinin farklı prevalansa sahip alanlarda hastalığın oluşumuna etkisini belirlemede geleneksel tüberkülozis analizlerine oranla karar ağaçlarının kullanılmasının daha başarılı sonuçlar verdiği ortaya konulmuştur.

Tyasi ve ark., (62) yaptıkları çalışmada; CART, CHAID ve kapsamlı CHAID algoritmaları ile karar ağacı modeli kullanılarak, tavukların morfolojik özelliklerine göre canlı ağırlıklarının tahmin edilebilirliğini araştırmıştır. Çalışmada; kanat uzunluğu, omurga uzunluğu, incik uzunluğu, göğüs çevresi, gaga uzunluğu, vücut uzunluğu, ayak

parmağı uzunluğu, vücut çevresi, sırt uzunluğu, incik çevresi ve kanat açıklığı ölçülerek veri setleri oluşturulmuştur. Daha sonra bu veriler; CART, CHAID ve kapsamlı CHAID algoritmaları ile karar ağacı modeli kullanılarak değerlendirilmiş ve karşılaştırılmıştır. Test edilen algoritmalar içerisinde CART algoritmasının canlı ağırlık tahmininde doğruluk oranı en yüksek algoritma olduğu belirlenmiştir. Çalışma sonunda CART algoritması kullanılarak en ağır tavukları üreten morfolojik özellikler belirlenmiştir. Bu özellikler gözetilerek tavuk yetiştiriciliği yapılmasının önerilebileceği belirtilmiştir.

Zhao ve Xue, (63), bazı miRNA'ların hedef tahmini için karar ağaçlarının kullanılabilirliğini araştırmışlardır. miRNA hedef belirleyicisi olarak; miranda (64), MİRDB (65) ve PITA isimli makine öğrenmesi ile ilişkili tahmin ediciler kullanılmıştır. Çalışma sonucunda bireysel tahmin edici uygulamalara oranla, çoklu eşikli karar ağaçlarının en az yüzde otuz daha başarılı olduğu belirlenmiştir.

Swain ve ark. (66), tek bir türü doğru olarak tanımlayabilmek için evrimsel olarak ilişkili birkaç türde kullanılabilir SNP barkodu dizilerinin üretilmesinde karar ağaçlarının kullanılabilirliğini araştırmışlardır. Bu amaçla; 64 adet *Anopheles* türü sivrisineğin mitokondriyal sitokrom oksidaz I (COI) geni analiz edilmiş ve toplam 32 adet SNP belirlenmiştir. Bu çalışmada ilk olarak; farklı *Anopheles* türlerinden 64 COI dizisi NCBI Gen Bankası kullanılarak belirlenmiştir. Belirlenen dizilerin uzunluklarındaki değişikliklerden dolayı belirlenen tüm COI dizileri, MEGA X yazılımında Clustal W uygulaması (67) kullanılarak hizalanmıştır. Ayrıca aynı dizi uzunluğunu sağlayabilmek için 5' ve 3' uçları kısaltılmıştır. İkinci aşama olarak; aynı uzunluktaki M nükleotidlerine sahip N türlerinin hizalanmış dizilerinden oluşan X destek verilerinin ifadesi formüle edilmiştir. Karar ağacı tabanlı barkodlama algoritması ile 64 *Anopheles* sivrisinek türünü ayırt edebilmek için 127 düğüm ve 32 lokus üretilmiştir. 64 tür için ilgili dizi etiketleri karar ağacı algoritmasına dayalı olarak tanımlanmıştır ve ağaç çizilmiştir. Karar ağacı algoritması ile üretilen türe özgü COI SNP'leri daha sonra bir boyutlu barkod desenlerine dönüştürülmüştür. Sonuç olarak karar ağacı algoritmasının *Anopheles* sivrisinek türlerinin ayırımı ve tanımlanması amacıyla kullanılan dizilerin etiketlerini oluşturmak için etkili bir yöntem olduğu ortaya konulmuştur.

#### 4. Sonuç

Veri madenciliğinin veteriner hekimliği alanında kullanımı, sunulacak hizmetlerin daha etkin olmasını ve kaynakların daha verimli kullanılmasını sağlaması yönünden önem arz etmektedir. Derlemenin konusu olan ve veri madenciliği yöntemlerinden biri olup karar verme mekanizmalarında öne çıkan karar ağaçları ivme kazanan bir şekilde uygulama alanı bulmaktadır.

Bu derleme ile değinilen bilgiler; karar ağaçlarının özellikle; veteriner hekimliği alanında teşhis bakımından farklı hastalık belirtileri ile hasta sahiplerine bilgi vermek, erken müdahale gerektiren durumlar ile tedavi değişiklikleri, kanatlı mortaliteleri düzeyleri, ıslah programları, mastitis gibi ekonomik önemi olan hastalıkların belirlenmesi, teşhis modeli geliştirme ve tedavi stratejilerinin etkin hale getirilmesi, genomik veriler kullanarak hayvanların kimlik teyidi çalışmaları, farklı parametreler açısından et kalitesi değerlendirmesi, miRNA'ların hedef tahmini, hayvan davranışları, tür tayini, ekonomik önemi olan verim özelliklerinin tahmini gibi birçok alanda kullanılabilirdiği ve kullanılabilirliği hususunda bilgi birikimine katkı sunmaktadır. Öyle ki; işaret edilen ve veteriner hekimliği alanında değerlendirilen daha çok teşhis koyma amacıyla kullanılan farklı karar ağacı algoritmaları günümüzde bilimsel arenada da üzerinde durulan konular olarak güncelliğini korumaktadır.

Buradan hareketle; bilimsel arenada ulaşılabilen literatürler özelinde değerlendirilen karar ağaçları ve veteriner hekimliği alanında karar ağaçlarının kullanımının konu hakkında yapılan çalışmalara temel dayanak oluşturacağı ve fikir vermesi açısından da anlamlı olacağı öngörülmüştür.



## Kaynaklar

1. Lovell MC. Data Mining. *Rev Econ Stat* 1983; 65(1): 1–12.
2. Coenen F. Data Mining: Past, Present and Future, *The Knowledge Engineering Review*, Cambridge University Press; 2004.
3. Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. *F Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press; 1996.
4. Ganesh SH. Data Mining: Should It Be Included In The 'Statistics' Curriculum? ICOTS6, 2002.
5. Koyuncugil A.S., Özgülbaş N. Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları, *Bilişim Teknolojileri Dergisi* 2009; 2(2): 21-32.
6. Ogunleye JO. The Concept of Data Mining. In: Thomas C, editor. *Data Mining - Concepts and Applications* [Internet]. London: IntechOpen; 2021.
7. Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: A literature review *J Biomed Inform* 2016; 61, 224-236.
8. Grzesiak W, Zaborski D Examples of the use of data mining methods in animal breeding. In: Karahoca A, editör *Data mining applications in engineering and medicine*, [Internet]. London: IntechOpen; 2012.
9. Hsiao T-C, Shieh S-L, Chen T-L, Liu C-H, Yeh Y-C. Data Analysis of Medical Records in Veterinary Hospital Using Clustering Method and Association Rule. *App Math Info Sci* 2015; 9(6):59.
10. Pandey SC. Data mining techniques for medical data: a review. 2016; SCOPES; 972-982.
11. Breiman L, Friedman JH, Stone CJ, Olshen RA. *Classification and regression trees*, 1st ed. New York: Chapman and Hall; 1984.
12. McCulloch W, Walter P. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull Math Biophys* 1943; 5 (4): 115–133.
13. Holland JH. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Michigan; 1975.
14. Duda RO, Hart PE. *Pattern Classification and Scene Analysis*. Wiley, New York; 1973.
15. Joseph B Application to the Logistic Function to Bio-Assay. *J Am Statist Assoc* 1944; 39: 357–365.
16. Han J, Kamber M. *Data mining: concepts and techniques*. Morgan Kaufmann, San Francisco, ABD; 2000.
17. Lewis RJ. An Introduction to Classification and Regression Tree (CART) Analysis. The 2000 Annual Meeting Of The Society For Academic Emergency Medicine, San Francisco, California, ABD; 2000.
18. Song Y, Lu Y. Decision tree methods: applications for classification and prediction. *Shangai Arch Psychiatry* 2015; 27(2):130-135.
19. Adua E, Kolog EA, Afrifa-Yamoah E, Amankwah B, Obirikorang C, Anto EO, Acheampong E, Wang W, Tetteh AY. Predictive model and feature importance for early detection of type II diabetes mellitus. *Transl Med Commun* 2021; 6:17.
20. Song Y, Lu Y. Decision tree methods: applications for classification and prediction. *Shangai Arch Psychiatry* 2015; 27(2):130-135.
21. Morgan JN, Sonquist JA. Problems in the analysis of survey data, and a proposal. *J Amer Statist Assoc* 1963; 58:415-434.
22. Kass GV. An explanatory technique for investigating large quantities of categorical data. *Appl Statist* 1980; 29(2):119-127.
23. Quinlan JR. Induction of decision trees. *Mach Learn* 1986; 1:81-106.
24. Biggs D, Ville BD, Suan E. A method of choosing multiway partitions for classification and decision trees. *J Appl Stat* 1991; 18:49-62.
25. Friedman J. Multivariate adaptive regression splines (with discussion), *Ann Stat* 1991; 19:1-141.
26. Quinlan JR. *C4.5: Programs for machine learning*, Morgan Kaufmann Publishers, 1993; 17-55.
27. Michie D, Spiegelhalter DJ, Taylor CC. *Machine learning, neural and statistical classification*. Ellis Horwood Limited, 1994.
28. Shafer J, Agarwal R, Mehta M. SPRINT: A scalable parallel classifier for data mining. *Proc. of 22nd International Conference on Very Large Databases*, 1996.
29. Loh W-Y, Shih Y-S. Split selection methods for classification trees. *Stat Sinica* 1997; 7:815-840.
30. Breiman L. Random Forests. *Mach Learn* 2001; 45: 5–32.
31. Friedman JH. Stochastic Gradient Boosting. *Comput Stat Data An*, 2002; 28:367-378.

32. Witten IH, Frank E, Hall MA. Data Mining Practical Machine Learning Tools and Techniques 3<sup>rd</sup> Edition, Morgan Kaufmann Publishers, Burlington, MA, ABD, 2011.
33. Rodriguez JJ, Kuncheva L, Alonso CJ. Rotation forest: A new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 2006; 28(10):1619-30.
34. Quinlan JR. Simplifying decision trees. *Int J Man-Mach Stud* 1987; 27: 221- 234.
35. Quinlan JR. Learning logical definition from relations. *Mach Learn* 1990; 5:239-266.
36. Quinlan JR. C4.5: programs for machine learning. Amsterdam: Elsevier, 2014.
37. Mingers J. An empirical comparison of pruning methods for decision tree induction. *Mach Learn* 1989; 4:227–243.
38. Webb GI. Overfitting. In: Sammut C, Webb GI, editors. *Encyclopedia of Machine Learning*. Springer, Boston, MA, ABD; 2011.
39. Vale MM, Moura DJ, Nääs IA, Oliveira SRM, Rodridues LHA. Data mining to estimate broiler mortality when exposed to heat wave. *Sci Agric* 2008; 65(3):223-229.
40. Piwczyński D. Using Classification Trees in Statistical Analysis of Discrete Sheep Reproduction Traits. *J Cent Europ Agri* 2009; 10 (3): 303-309.
41. Pinzón-Sánchez C, Cabrera VE, Ruegg PL. Decision tree analysis of treatment strategies for mild and moderate cases of clinical mastitis occurring in early lactation. *J Dairy Sci* 2011; 94:1873-1892.
42. Židek R, Šidlová V, Kasarda R, Fuerst-Waltl B. Methods for Distinction of Cattle Using Supervised Learning. *Int J Biol Vet Agri Food Engin* 2014; 8(5):500-502.
43. Wylie CE, Shaw DJ, Verheven KLP, Newton JR. Decision tree analysis of clinical data to aid diagnostic reasoning for equine laminitis. *Vet Record* 2016; 178(17):420.
44. Velásquez L, Cruz-Tirado JP, Siche R, Quevedo R. An application based on the decision tree. to classify the marbling of beef by hyperspectral imaging. *Meat Science* 2017; 133:43-50.
45. Wold JP, Kermit M, Woll A. Rapid nondestructive determination of edible meat content in crabs (*Cancer pagurus*) by near-infrared imaging spectroscopy. *Applied Spectroscopy* 2010; 64(7): 691-699.
46. López Osornio MM, Hough G, Salvador A, Chambers IV E, McGray S, Fiszman S. Beef's optimum internal cooking temperature as seen by consumers from different countries using survival analysis statistics. *Food Qual Prefer* (2008; 19(1):12-20.
47. Gagaoua M, Monteils V, Picard B. Decision tree, a learning tool for the prediction of beef tenderness using rearing factors and carcass characteristics. *J Sci Food* 2019; 99:1275-1283.
48. Ekiz B, Baygül O, Yalçın H, Özcan M. Comparison of the decision tree, artificial neural network and multiple regression methods for prediction of carcass tissues composition of goat kids. *Meat Sci* 2020; 161:108011.
49. Tamura T, Okubo Y, Deguchi Y, Koshikawa S, Takahashi M, Chida Y, Okada K. Dairy cattle behavior classifications based on decision tree learning using 3 axis neck mounted accelerometers. *Anim Sci J* 2019; 90:589-596.
50. Piwczyński D, Sitowska B, Kolenda M, Brzozowski M, Aerts J, Sshork PM. Forecasting the milk yield of cows on farms equipped with automatic milking system with the use of decision. trees. *Anim Sci J* 2020; 91:e13414.
51. Pascottini OB, Probo M, Leblanc SJ, Opsomer G, Hostens M. Assessment of associations between transition diseases and reproductive performance of dairy cows using survival analysis and decision tree algorithms. *Prevent Vet Med* 2020; 176:104908.
52. Kamphuis C, Mollenhorst H, Feelders A, Pietersma D, Hogeveen H. Decision-tree induction to detect clinical mastitis with automatic milking. *Comput Electron Agric* 2010a; 70:60–68.
53. Firk R, Stamer E, Junge W, Krieter J. Improving oestrus detection by combination of activity measurements with information about previous oestrus cases. *Livest Prod Sci* 2003; 82:97-103.
54. Sun Z, Samarasighe S, Jago J. Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks *J Dairy Res* 2010; 77:168-175.
55. Caraviello DZ, Weige KA, Craven M, Gianola D, Cook NB, Norlund KV, Fricke PM, Wiltbank MC. Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. *J Dairy Sci* 2006; 89:4703-4722.
56. Shahinfar S, Page D, Guenther J, Cabrera V, Fricke P, Wigel K. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *J Dairy Sci* 2014; 97:731-742.

- 
57. Romero MP, Chang Y-M, Brunton LA, Parry J, Prosser A, Upton P, Rees E, Tearne O, Arnold M, Stevens K, Drewe JA. Decision tree machine learning applied to bovine tuberculosis risk factors to aid disease control decision making. *PreventVet Med* 2020; 175:104860.
  58. Hilbe JM. *Logistic Regression Models*, 1st ed. Chapman & Hall/CRC; 2009.
  59. Campbell MJ, Swinscow TDV. *Statistics at Square One*, 11th ed. BMJ Publishing Group Ltd, UK; 2009.
  60. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw* 2008; 28.
  61. Therneau, T, Atkinson E. *An Introduction to Recursive Partitioning Using the Rpart Routines*. Boca Raton, Florida: Chapman; Hall/CRC; 2019.
  62. Tyasi TL, Eyduran E, Celik S. Comparison of tree-based regression tree methods for predicting live body weight from morphological traits in Hy-line silver brown commercial layer and indigenous Potchefstroom Koekoek breeds raised in South Africa. *Trop Anim Health and Pro* 2021; 53:7
  63. Zhao B, Xue B. Improving prediction accuracy using decision tree based meta strategy and multi-threshold sequential voting exemplified by miRNA target prediction. *Genomics* 2017; 109:227-232.
  64. Enright AJ, John EB, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in drosophila, *Genome Biol* 2003; 5:R1.
  65. Wang X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics* 2016; 32:1316–1322.
  66. Swain SN, Makunin A, Simanchal Dora A, Barik TK. SNP barcoding based on decision tree algorithm: A new tool for identification of mosquito species with special reference to Anopheles. *Acta Tropica* 2019; 199: 105152.
  67. Kumar S, Stecher G, Li M, Knyaz M, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018; 35: 1547–1549.