



Hayvancılıkta Veri Madenciliği Uygulamaları *

Feza ALEV ÇETİN, Nazire MİKAIL **

Siirt Üniversitesi, Ziraat Fakültesi, Zootečni Bölümü, Siirt, TÜRKİYE

Geliş Tarihi/Received: 14.10.2015

Kabul Tarihi/Accepted: 27.01.2016

**Sorumlu yazar/Corresponding author: naziremikail@siirt.edu.tr

Özet: Veri madenciliği bir merkezde toplanan çok büyük miktardaki bilgilerden istenilen ve uygulanabilir bilginin keşfedilmesini sağlamaktır. Veri madenciliği bilgi endüstrisi ve toplumunda kullanılmaya başlanmıştır. Pek çok veri madenciliği metodu kullanılmakla birlikte, son yıllarda bu teknikler hayvancılık alanında dikkat çekici olmuştur. Hayvancılıkla ilgili karmaşık problemlerin çözümü için pek çok metot ele alınmış ve geliştirilmiştir. Çalışmada k-ortalamları 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 veri madenciliği yöntemleri hakkında kısa bilgi verilmiştir. Bu çalışma ile veri madenciliği metotları tanıtılacak ve dünyada veri madenciliğinin hayvancılık alanındaki uygulamalarına örnekler verilecektir.

Anahtar Kelimeler: Hayvancılık, veri madenciliği, yapay sinir ağları, destek vektör makineleri, karar ağaçları

Data Mining Applications in Livestock

Abstract: Data mining provides discovering the required and applicable knowledge from very large amounts of information collected in one centre. Data mining has been used in the information industry and society. Although many methods of data mining has been used, these techniques has been remarkable in animal husbandry in recent years. For the solution of complex problems in animal husbandry many methods were discussed and developed. Brief information on data mining techniques such as k-means approach, k-nearest neighbor approach, multivariate adaptive regression function (MARS), naive Bayesian classifiers (NBC), artificial neural networks (ANN), support vector machines (SVM), decision trees are given in the study. Some data mining methods are presented and examples of the application of data mining in the field of animal husbandry in the world are provided with this study.

Keywords: Livestock, data mining, artificial neural networks, support vector machines, decision trees

1. Giriş

Veri madenciliği bilgi teknolojilerinin doğal bir süreci olarak görülebilir. Elde edilen fazla miktarda bilgiyi depolamayı, uzun yıllar sürecektir uygulamalar için yorum çıkartabilmeyi ve ileriye dönük tahminlemeler yapabilmeyi sağlayan bir bilgi sürecidir. Bu süreç verinin temizlenmesi, verilerin birleştirilmesi, istenilen verinin seçimi, verileri uygun formlara dönüştürme, veri madenciliği uygulamaları ve yorumlanması şeklinde sıralanır (Rajesh, 2011).

Literatürde yer alan çok farklı veri madenciliği metotları vardır. Bunlar arasında regresyon ve sınıflandırma ağaçlarının genel modelleri, ki-kare otomatik interaksyon belirleme modelleri (Chi-Squared Automatic Interaction Detection, CHAID), etkileşimli sınıflandırma ve regresyon ağaçları, çok değişkenli uyarlanırlı regresyon eğrileri (Multivariate Adaptive Splines, MARS), yapay sinir ağları (Artificial Neural Networks, ANN), destek vektör makineleri (Support Vector Machines, SVM), k en yakın komşuluklar (k Nearest Neighborhoods, k NN), k-ortalamları

*9. Ulusal Zootečni Bilim Kongresinde poster olarak sunulmuş ve özet olarak basılmıştır.

yaklaşımı, Bayes sınıflandırıcıları (Naive Bayesian Classifiers, NBC) gibi metotlar sayılabilir. Bu metotlar hayvan ıslahı ve yetiştiriciliği ile ilgili pek çok konuda uygulanabilmektedir (Grzesiak ve Zaborski, 2012).

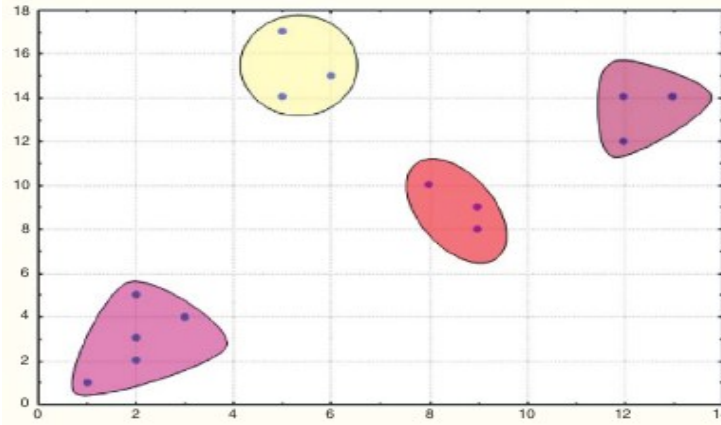
Veri madenciliği teknikleri; sınıflandırma teknikleri, kümeleme teknikleri ve birliktelik kuralı bulma olarak başlıca üç gruba ayrılır (Kahramanlı, 2008). Sınıflandırma teknikleri, sınıflandırma yapılmış bir setin içinden sağlanan bilgi ile sınıfı bilinmeyen örnekleri sınıflandırmak için kullanılır. Şayet sınıflandırılmış bir veri yoksa bu durumda kümeler ayrılma şeklinde kümeleme teknikleri kullanılır. Sınıflandırma yapılan verilerde genellikle bir eğitim seti kullanılır. Bunlara yapay sinir ağları ve destek vektör makineleri iyi bir örnek olup, en çok kullanılan kümeleme tekniklerinden biri de k-ortalamları yaklaşımıdır (Mucherino ve ark., 2009).

Bu çalışmada, veri madenciliği metotlarından bazıları incelenerek, hayvancılık alanında veri madenciliği uygulamalarına yer verilmiştir.

2. Veri Madenciliği Metotları

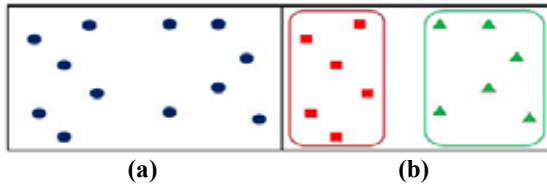
2.1. K-ortalamları yaklaşımı

Sınıfı bilinmeyen örnekleri sınıflandırırken belli bir eğitim seti yoksa K-ortalamları yaklaşımı veri madenciliğinde kümeleme tekniği için uygun bir metottur (Hartigan, 1975). Bu teknikte amaç, kümeler içinde yer alacak benzer verileri gruplara ayırmaktır. Her kümenin içindeki örnekler kümenin merkezine eşit uzaklıktadır ve kümenin merkezi küme içerisindeki tüm örneklerin ortalamasıdır. Bir küme içerisindeki örnek başka bir kümenin merkezine de yakın olduğu durumda, örnek kendi kümesinden yeni bir kümeye taşınır (Bishop, 2006). K-ortalama algoritma yapısı Şekil 1'de gösterilmiştir.



Şekil 1. Kümeleme probleminde veri kümeleri (Nisbet ve ark., 2009)

K-ortalamları algoritması sadece kümelerin ortalamalarına göre uygulama yapar. Kümeler oluşturulurken kümelerin kovaryanslarını tahmin etmemesi dikkate alınmalıdır (Bishop, 2006). Şekil 2'de bir Kartezyen uzayda elde edilen örnek verilerle ortalamalar dikkate alınarak oluşturulan kümeler gösterilmiştir.

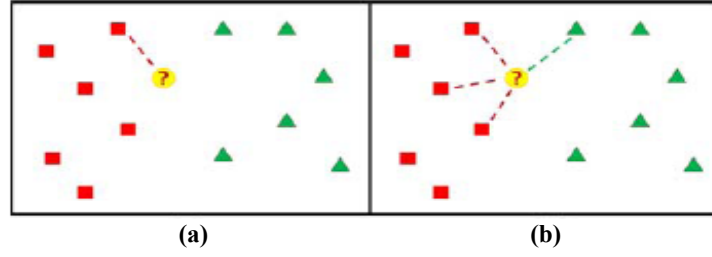


Şekil 2. Bir Kartezyen uzayda noktalardan oluşan bir verinin en uygun kümeler ayrılması: a) noktalar henüz kümelere bölünmemiştir, b) aynı kümeye ait noktalar aynı sembol kullanılarak işaretlenmiştir (Bishop, 2006)

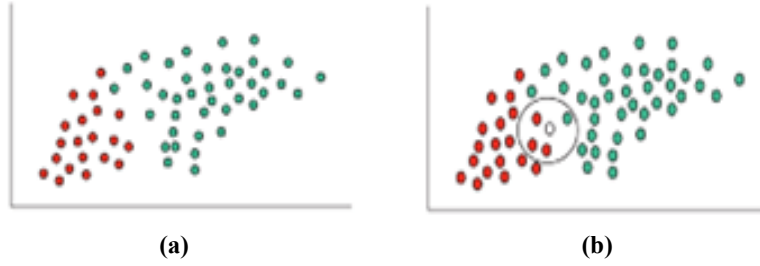
2.2. K-en yakın komşu yaklaşımı

K-en yakın komşu yaklaşımı sınıflandırma için uygulanan bir tekniktir. Sistem işleyişi için bir eğitim verisi mevcuttur. Sınıfı bilinmeyen örnekler bu eğitim verisine göre sınıflandırılmaktadır. Bu yaklaşımın temelinde benzer örneklerin aynı sınıfta olması gerektiğinden, örnekler arası benzerlikler uygun uzaklık fonksiyonları kullanılarak ölçülmelidir (Şekil 3) (Mucherino ve ark., 2009).

Burada k parametresi benzer bilinen örneklerin sayısını verir. Bilinmeyen bir örnek verildiği zaman, onun eğitim verisindeki tüm örneklerden uzaklığı hesaplanır. K-en yakın bilinen örnekler bu hesaplamaya göre yerleştirilir, ardından sınıfı bilinmeyen yeni veriler için bilinen örnekler arasındaki en sık kullanılan sınıflandırmaya göre değerlendirilir (Şekil 4).



Şekil 3. “?” sembolü ile işaretli nokta onun en yakın komşu sınıflarına göre sınıflandırılmıştır: a) $k=1$ ve bilinmeyen nokta kareler tarafından işaretli sınıfa ait olarak sınıflandırılmıştır, b) $k=4$ ve bilinmeyen noktada kareler tarafından işaretli sınıfa ait olarak sınıflandırılmıştır (Mucherino ve ark., 2009)



Şekil 4. (a) iki eksene tanımlanmış bir analiz uzayında nesnelere kırmızı ve yeşil olarak iki grupta sınıflandırılmıştır, (b) yeni nesnelere (beyaz top) analiz uzayındaki pozisyonu gösterilmiştir (Nisbet ve ark., 2009)

2.3. Çok değişkenli uyarlanırlı regresyon eğrileri (MARS)

Çok değişkenli uyarlanırlı regresyon eğrileri regresyon tipi modelleri çözmek için kullanılan bir yöntem olup, regresyon dışında, sınıflandırma için de kullanılabilir. Sadece iki durumda, bağımlı değişken ikili olarak kodlanır ve diğer işlemler regresyon problemleri ile aynı işler. Uygun bir MARS modeli iki aşamadan oluşur. İlk aşamada, eğrinin uyum sağlaması için birçok temel fonksiyon içeren model oluşturmaktır. Bu aşamada, ya ilave bileşenler oluşturulabilir ya da tahminler arasındaki interaksyonları hesaplamak mümkün olabilir (Lee ve ark., 2006). Uygulama algoritmasının ikinci aşamasında, uyum derecesi için katkı sağlayan temel fonksiyonlar uzaklaştırılır (Zareipour ve ark., 2006). Çok değişkenli uyarlanırlı regresyon eğrilerinin özü, bağımlı değişken ve tahminler arasındaki ilişkinin tahmini için temel fonksiyonların birleşiminden oluşmasıdır (Grzesiak ve Zaborski, 2012).

2.4. Bayes sınıflandırıcıları (NBC)

Bayes sınıflandırıcıları koşullu olasılık durumunda Bayes'in teoreminde kullanılan bir tekniktir. Özel bir konu için konunun değerlendirilmesi, en çok olası sınıf, veri tabanlı tahmin edilen ihtimallerin bir ürünü olan olasılık

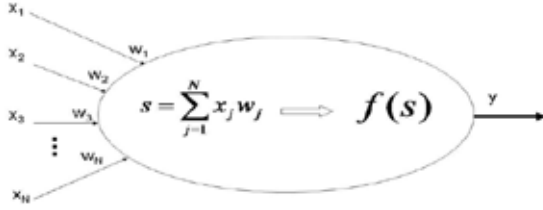
hesaplanması yoluyla yapılır (Grzesiak ve ark., 2011). Bu ihtimaller sınıflandırılmış konunun değişkenlerinin değerlerine bağlı seçilir (Larose, 2006). Tanımlanan değişkenler hem kesikli (örneğin, laktasyon sayısı), hem de sürekli (örneğin, süt verimi) olabilir. Bu sınıflandırma östrus döngü fazının sınıflandırılması için hayvan yetiştiriciliğinde kullanılmıştır (Maldonado-Castillo ve ark., 2007).

2.5. Yapay sinir ağları (ANN)

Yapay sinir ağları insan beynine benzeyen biyolojik sistemlerden etkilenmiş canlı bir sinir hücresinin basitleştirilerek modellenmiş şeklindedir. Yapay sinir ağları, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır. Genel olarak yapay sinir ağları, model seçimi ve sınıflandırılması, işlev tahmini, en uygun değeri bulma ve veri sınıflandırması gibi işlerde başarılıdır (Elmas, 2011).

Yapay sinir ağlarında her düğüm bir nöronu temsil eder ve her bağlantı iki nöron arası etkileşimi gösteren bir ağ yapısıdır (Şekil 5). Yapay sinir ağlarının en çok kullanılan çeşidi çok katmanlı algılayıcıdır ki, burada nöronlar tabakalarda organize olmuşlardır (Mucherino ve

ark., 2009). Bir tabakanın nöronları sonraki tabakanın tüm nöronlarıyla bağlantılıdır. Çoğu zaman yapay sinir ağları üç katmandan oluşmuş olur. İlk katman giriş katmanıdır ve burada girişler ağa tanıtılır. Diğer katman gizli katman veya katmanlardır ve burada girişler işlenir. Son katman da çıkış katmanıdır. Burada verilmiş girişlere uygun çıkışlar üretilir (Mammadova, 2012).



Şekil 5. Yapay bir nöronun şematik gösterimi (Grzesiak ve Zaborski, 2012)

S= Postsinaptik potansiyel, $X_j=j$. girdi sinyali, $W_j= j$. girdi nöronu ile ilişkili ağırlık, $N=$ Girdi nöronlarının sayısıdır.

2.6. Destek vektör makineleri (SVM)

Destek vektör makineleri metodu makine öğrenmesinin istatistik alanı üzerine kurulmuş bir uygulamasıdır. Destek vektör makineleri orijinal olarak kendine has formüllerle düzlemi iki ayrı sınıfa direkt olarak ayırabilir. Destek vektör makineleri; öğrenme, sınıflandırma, kümeleme, yoğunluk tahmini ve son olarak da veriden regresyon kuralları üretmek için kullanılan bir eğitime algoritmasıdır. Destek vektör makinelerinde amaç, veri noktalarını mümkün olduğu kadar iyi sınıflandıran ve iki sınıf noktaya ayıran optimum ayırıcı düzlemin bulunmasıdır. Diğer bir ifadeyle, iki sınıf arasındaki uzaklığın maksimum olduğu durumun bulunması amaçlanmaktadır. Bu sınıflandırma mantığının temel taşları ise, her iki sınıfın uç noktalarında bulunan ve eğitime örneklerinin arasından seçilen destek vektörleridir (Mammadova, 2012).

Destek vektör makineleri iki ayrı sınıftaki veri örneklerini sınıflandıran, ikili sınıflandırıcılardır (Cortes ve Vapnik, 1995). Bu tekniğin temeli, iki sınıfın doğrusal olarak ayrıldığı basitleştirilmiş bir yöntemdir (Şekil 6).

2.7. Karar ağaçları

Karar ağaçları veri madenciliğinde yaygın şekilde uygulama alanı bulmuştur. Sınıflandırma ağaçları veri madenciliği grubu araçlarından olup, modern analitik bir tekniktir (Abu-Hanna ve Keizer, 2003). Karar ağaçları hem tahmin hem de tanımlama için kullanılabilen grafiksel modellerdir. Sınıflandırma ve regresyon ağaçları (Classification and Regression Trees, CART) karar

ağaçlarının bir çeşididir (Grzesiak ve Zaborski, 2012).

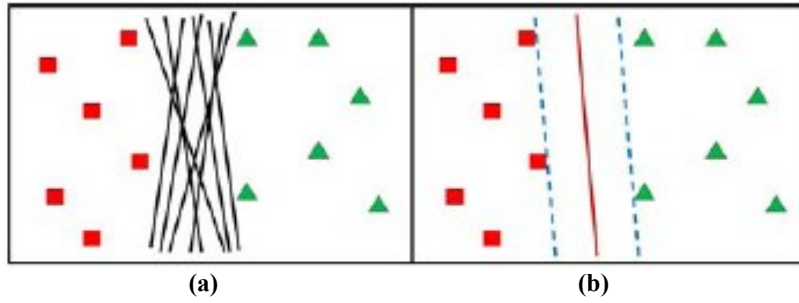
Sınıflandırma ve regresyon ağaçlarının karakteristik özelliği karar ağaçlarının tam anlamıyla ikili algoritma yapısıdır. Eğitim verileri hedef değişkenlerin benzer değerlerine sahip olan alt kümelere ardışık olarak bölünür ve ağaç her karar düğümü için mevcut tüm değişkenlerin ve tüm ihtimal bölünmelerin kapsamlı araştırılmasıyla ve verilen kritere göre en uygun kümelerin seçimi yoluyla inşa edilir (Şekil 7).

Karar ağaçları kök düğümlerinden uç yaprak düğümlerine giden dallarla bir araya gelen karar düğümleri kümesidir (Piwczyński, 2009). Sınıflandırma ve regresyon ağaçlarında sadece iki alt düğüm bir karar düğümünden doğrudan dallanabilir. Ağaç eğitim veri kümesinin tekrarlı bölümleri ile kurulur. Her aşamada kayıtlar homojen olarak iki alt kümeye bölünür. Her alt kümede aynı bölünme süreci tekrarlanır (Grzesiak ve ark., 2011). Ayrıca, ağacın büyüklüğü bölünen alt kümelerin toplam sayısına bağlıdır (De'ath ve Fabricius, 2000).

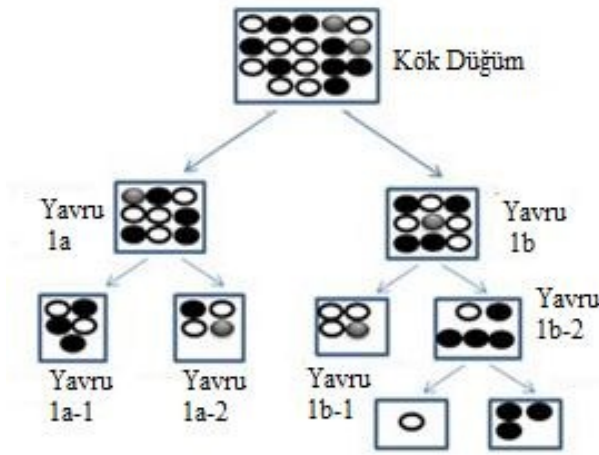
3. Hayvancılıkta Veri Madenciliği Uygulamaları

Veri madenciliğinde çeşitli teknikler kullanılmakla birlikte, bu tekniklerle yapılmış hayvancılık alanındaki bazı çalışmalar şu şekildedir:

Grzesiak ve ark. (2010), suni tohumlamada sorunla karşılaşılacak ineklerin belirlenmesi ve potansiyel risk olabileceklerin elenmesi için yapay sinir ağları ve çok değişkenli uyarlanırs regresyon eğrileri metotlarını uygulamışlardır. İnekler bir ya da iki çiftleştirmeden sonra gebe kalırlarsa “iyi”, gebelik için iki çiftleştirmeden fazlasına ihtiyaç varsa “kötü” olarak iki gruba ayrılmıştır. Bu çalışmada sınıflandırma fonksiyonu ve lojistik regresyon oldukça düşük sonuç gösterirken, en iyi performansı yapay sinir ağları ve çok değişkenli uyarlanırs regresyon eğrileri göstermiştir. Suni tohumlamada probleme sahip inekleri belirlemek için belirli değişkenler kullanılmıştır. Her iki modelde de buzağılama aralığı en önemli belirleme değişkeni olarak bulunmuştur. Yapay sinir ağları tarafından gösterilen sonraki değişkenler sırasıyla laktasyon sayısı, vücut durum skoru, gebelik süresi ve inek genotiplerindeki Holstein-Friesian genlerinin yüzdesidir. Çok değişkenli uyarlanırs regresyon eğrileri modelinde ise, buzağılama aralığından sonra en büyük katkıyı vücut durum skoru ve gebelik süresinin sağladığı bulunmuştur. Ayrıca çok değişkenli uyarlanırs regresyon eğrileri modeli yaş, süt verimi, süt yağı, protein içeriği ve laktasyon sayısını da içermiştir. “Kötü” olarak



Şekil 6. Kartezyen sistemdeki noktalar onların özelliklerine göre ayrılır ve iki farklı sınıfta değerlendirilir: a) maksimize edilmemiş hiperdüzlemde iki sınıf arasındaki ayrılma ihtimalleri, b) hiper düzlemde ayırım en yüksek marjın destek vektör makineleri tarafından bulunmuştur (Mucherino ve ark., 2009)



Şekil 7. Örnek bir karar ağacı yapısı (Nisbet ve ark.,2009)

değerlendirilen tohumlamada sorunlu inekleri belirlemede çok tabakalı algılayıcı sinir ağı (Multilayer Perceptron, MLP) % 87 ile en yüksek hassaslık göstermiştir. Çok değişkenli uyarlanırların regresyon eğrilerinin hassaslığı ise % 86 olmuştur. Belirlilik çok tabakalı algılayıcı sinir ağı ve çok değişkenli uyarlanırların regresyon eğrileri için % 85 olarak bulunmuştur. Yapay sinir ağı ve çok değişkenli uyarlanırların regresyon eğrisinin geçerliliğinin diğer istatistik metodlarından daha iyi olduğu belirlenmiştir.

Grzesiak ve ark. (2011) yaptıkları çalışmada veri madenciliği metodlarından yararlanarak süt ineklerindeki gebe kalma farklılıklarının belirlenmesi için NBC ve regresyon sınıflandırma ağaçları kullanmışlardır. Veri olarak; laktasyon sayısı, yapay tohumlama mevsimi, inek genotiplerindeki Holstein-Friesian genlerinin oranı, tohumlanmış ineğin yaşı, bir önceki buzağılamadaki buzağının cinsiyeti, önceki gebelik ile buzağılama aralığı, bir önceki gebelikten bir sonraki buzağılama sonuna kadarki gün sayısı, gebelik süresi, inek vücut durum endeksi, sütteki

yağ oranı, sütteki yağ ve protein içeriği yüzdesi olmak üzere 11 tane değişkeni kullanılmıştır. İnekler, bir ya da iki çiftleştirmeden sonra gebe kaldıysa "iyi", gebelik için iki çiftleştirmeden fazlasına ihtiyaç varsa "kötü" olarak iki gruba ayrılmıştır. Bu çalışmada belirlilik her iki sınıf için de benzer bulunmuştur (sınıflandırma ve regresyon ağaçları için % 90 ve Bayes sınıflandırıcıları için % 93). Bayes sınıflandırıcıları ile daha yüksek belirlilik olmasına rağmen, Bayes sınıflandırıcıları için hassaslık istatistiksel olarak sınıflandırma ve regresyon ağaçlarından daha düşük çıkmıştır. Bayes sınıflandırıcıları ile sınıflandırma ve regresyon ağaçları metodları bu çalışmada karşılaştırıldığında, doğumda zorlukla karşılaşan süt ineklerinin belirlenmesinde sınıflandırma ve regresyon ağaçlarının yetiştiriciler için daha avantajlı olduğu ve suni tohumlamada herhangi bir sıkıntısı olmayan ineklerde de daha hassas olduğu vurgulanmıştır. Kullanılan değişkenler arasında vücut durum skor endeksi, sınıflandırma ve regresyon ağaçları metodunda önemli rol oynamıştır. Bu çalışmada, sınıflandırma ve

regresyon ağaçlarının kullanımı ile; sürüde doğurganlığı ve yetiştiricilerin ekonomik etkinliğinin devamlılığını arttırmak için, zor gebe kalan ineklerde doğrudan koruyucu yönde destek sağlanabildiği vurgulanmıştır.

Kamphuis ve ark. (2010) yaptıkları çalışmada, otomatik olarak sağım yapılan hayvanlarda klinik mastitis oluşumunun varlığını göstermek için karar ağaçları modellerini kullanmışlardır. Veriler otomatik sağım yapan ticari çiftliklerden toplanmıştır. Bu çalışmada karar ağaçlarının oluşturulması uygulamaya dönük olmamakla birlikte, klinik mastitisin belirlenmesi için oluşturulacak karar ağacı modellerinin gelişimi için kanıt olduğu ve bu potansiyelin varlığının bu çalışma ile doğrulandığı vurgulanmıştır.

Vale ve ark. (2008), çalışmalarında ısı dalgasına maruz kalan broyler tavuklarının ölümlerini tahmin etmek için veri madenciliği metodlarından karar ağaçlarını kullanmışlardır. Çevresel veri olarak sıcaklık ve nem endeksi (Temperative Humidity Index, THI) kullanılmıştır. Meteorolojik veriler ve meteoroloji istasyonlarından hesaplanan sıcaklık ve nem endeksi, 29 ve 42 günlük broylerler için zararlı çevre koşullarının aralığını belirlemek için yararlı olmuştur. Modeli kurmak için kullanılan veri tabanı; broyler ölümü, ilgili barınaklar ve dış çevre verileri (kuru termometre sıcaklığı, nemli termometre sıcaklığı, rüzgâr hızı, maksimum ve minimum sıcaklıklar, nispi nem, hesaplanan sıcaklık ve nemlilik endeksi) verileri ile oluşturulmuştur. Veri madenciliği teknikleri 3 modelin gelişmesine olanak sağlamış ve bu modellerden temel bileşen analizi ve Wrapper özellik seçimi yaklaşımı % 89.3 doğrulukla broyler ölümlerini tahmin etmiştir. Bu çalışma ile karar ağaçlarının broyler tavuk ölümlerini tahmin ederken veri sınıflandırmasında kullanılabileceği ortaya konulmuştur.

Piwczyński (2009), çalışmasında sınıflandırma ağaçlarını anaç koyunlar tarafından büyütülen yavruların sayısındaki değişimden sorumlu faktörleri belirlemek için kullanmıştır. Anaç koyun tarafından yetiştirilen yavru sayısındaki çeşitlilik; 70 gün ve 12 aylık olan koyun vücut ağırlığı, üretim sektörüne (kamu, özel), koyun ve ebeveynlerin doğum tipi, koyun yetiştirme tipi, koyun doğum yılı (1985-1998), kuzulama yılı (1993-2000), koyun yaşı (2-8 yıl arası) ve sürüye göre değerlendirilmiştir. Karar ağaçları tekniği yetiştirilen yavru sayısını etkileyen; koyun yaşı, sürü, doğum tipi ve yavrunun 12. ay canlı ağırlığı bağımsız değişkenleri ile tanımlanmıştır. Bu çalışmada, sürünün yetiştirilen yavru sayısını etkileyen çok önemli bir faktör olduğu ve veri alt

kümelerini oluşturan faktörün koyun vücut ağırlığı olduğu belirlenmiştir. Araştırmada ayrıca, 12 aylıkken vücut ağırlığının etkisinin ıslah bakımından daha önemli olabileceği vurgulanmıştır.

Zaborski ve Grzesiak (2011a)'ın yaptıkları çalışmada ER α -BglI, ER α -SnaBI ve CYP19-PvuII genotipleri ile ilgili yapay sinir ağları kullanılarak sürülerde distosinin (zorlu doğum) belirlenmesine çalışılmıştır. Araştırmacılar girdi değişkeni olarak; düve genotiplerinde Holstein-Friesian genlerinin yüzdesi, gebelik süresi, vücut durum skoru, buzağılama mevsimi, buzağılama yaşı ve önceden seçilmiş üç genotip kullanmışlardır. İkiye ayrılmış çıktı değişkeni doğum zorluğuna göre kolay ve zor olarak sınıflandırılmıştır. Bu çalışmanın amacı; seçilen tanısıl değişkenler baz alınarak düvelerdeki distosinin belirlenmesi için kullanılan yapay sinir ağı tipinin hazırlık ve kalitesinin değerlendirilmesi, bu belirlemenin uygulanması ve varolan distosideki en büyük etkiye sahip değişkenleri bulmaktır. Araştırmada; kolay ve zor doğum olarak sınıflandırmak için düvelerde yapay sinir ağlarının farklı tipleri uygulandığında, gizli tabakada 4 nöronlu çok tabakalı algılayıcı sinir ağları distosiyi belirlemede en iyi sonuç göstermiştir. En etkili faktörler; gebelik süresi, doğumda ve doğumdan önce vücut durum skor endeksleri arasındaki farklılık, düve genotiplerinde Holstein-Friesian genlerinin yüzdesi ve ER α -BglI, CYP19-PvuII genotipleridir. Doğum sırasında ve doğumdan önce vücut durum skoru arasındaki fark doğum sınıflandırmasını etkilediği için önemli bir değişkendir. Açıklayıcı değişkenlerin bir kısmı yapay sinir ağları tarafından kullanılarak düvelerdeki distosinin belirlenme doğruluğu oldukça yüksek seviyede bulunmuş ve yapay sinir ağları düvelerde distosinin önlenmesi için dolaylı bir araç olarak kullanılabileceği belirtilmiştir.

Zaborski ve Grzesiak (2011b), süt ineklerinde distosiyeye (zorlu doğum) sahip olanları yapay sinir ağları kullanarak belirlemeye çalışmışlardır. Çalışmanın amacı; süt ineklerinden zorlu doğum yapanları belirlemek için seçilen yapay sinir ağı modelinin kalitesini değerlendirmek ve hazırlamak, seçilen modelin buzağılama ile ilgili probleme sahip hayvanları belirlemesinin değerlendirilmesi, doğum sürecinde en büyük etkiye sahip faktörü bulmaktır. Yapay sinir ağları tarafından bu belirlemenin doğruluk seviyesi oldukça yüksek bulunmuştur. Zorlu doğuma sahip süt ineklerinin belirlenmesine en büyük katkıyı sağlayan değişkenlerin buzağılama mevsimi, CYP19-PvuII genotipi, gebelik süresi, vücut durum skor endeksi, inek genotiplerindeki Holstein-Friesian genlerinin oranı, laktasyonda en

yüksek (pik) süt verimi ve yaşın olduğu gösterilmiştir.

Grzesiak ve ark. (2003), Holstein-Friesian ineklerindeki 305 günlük laktasyon süt veriminin tahmininde kullanılan yapay sinir ağları ve çoklu regresyon tahminlerini karşılaştırmışlardır. Çalışmada; barınaktaki 305 günlük süt verimi ortalaması, sağımda geçen gün sayısı, araştırma periyodunun 1., 2., 3., 4. ayındaki kontrol günü süt verim ortalaması ve buzağılama ayı olmak üzere 7 girdi değişkeni kullanılmıştır. Tahminleme hem yapay sinir ağları, hem de çoklu regresyon modeli kullanılarak yapılmıştır. Polonya süt sığırlarının değerlendirme sistemlerinin bulguları ile tahminleme sonucunda alınan değerler farklı çıkmamıştır. Bu yüzden yapay sinir ağları modeli tahminleme metodlarına alternatif olduğu vurgulanmıştır. Grzesiak ve ark. (2006)'da benzer bir çalışma ile süt ineklerinde süt verim tahminlerini yapay sinir ağları ve Wood'un laktasyon eğrisi metodu (Wood, 1967) ile karşılaştırarak uygulamışlardır. Bu çalışma sonucunda yapay sinir ağlarının kalite parametreleri Wood'un tahminleme modeli ile kıyaslamada daha iyi bulunmuştur. Yapay sinir ağlarının parametreleri daha düşük tahminleme hatasına sahip olduğu görülmüştür. Çalışmada yapay sinir ağlarının kullanımı regresyon modellerinin varsayımlarıyla karşılaştırma gerektirmeyeceği ve regresyon modelinden daha kolay kullanıma sahip olduğuna değinilmiştir.

Benzer bir çalışma Görgülü (2012) tarafından, İsviçre Esmeri sığırlarında 305 günlük süt verimi tahmini için yapay sinir ağları kullanılarak yapılmıştır. Yapay sinir ağları çoklu lineer regresyonlarla karşılaştırılarak geliştirilmiştir. Her inek; 305 günlük süt verimi ortalaması, yaş, laktasyon sayısı, buzağılama mevsimi, örneklerin sırasıyla 1., 2., 3., 4. aydaki kontrol günü süt verimi ortalamaları olmak üzere 8 değişkenle tanımlanmıştır. Bu çalışmada yapay sinir ağları modeli değişkenlerden özellikle 1., 2., 3. ve 4. kontrol günü kayıtları ile 305 günlük süt verimi için en iyi tahmini sağladığı belirtilmiştir.

Takma ve ark. (2012) yaptıkları benzer bir çalışmada Siyah Alaca ineklerin laktasyon süt verimleri üzerine laktasyon süresi (LS), buzağılama yılı (BY) ve servis periyodunun (SP) etkisini çoklu regresyon ve yapay sinir ağı modelleri ile araştırmış ve uyum yeteneklerini karşılaştırmıştır. Yapay sinir ağı ve çoklu regresyon modelleri karşılaştırıldığında, süt verimlerinin tahminlenmesinde yapay sinir ağı modelinin çoklu doğrusal regresyon modelinden daha iyi uyum sağladığı gözlenilmiştir. Yapay sinir ağlarının çoklu doğrusal regresyon analizine göre

daha yüksek bir öngörüye sahip olduğu ve daha az hataya sahip sonuçlar verdiği belirtilmiştir. Bunun sonucunda yapay sinir ağlarının regresyon analizine alternatif bir metot olabileceği sonucuna varılmıştır.

Mammadova ve Keskin (2013) çalışmalarında, süt ineklerinde klinik ve subklinik mastitis tahmini için destek vektör makinelerini uygulamışlardır. Özel işletmede yetiştirilen Holstein ineklerden 12 ay boyunca ayda bir gün olmak üzere elde edilen sütteki somatik hücreler sayılarak, hayvanlar somatik hücre sayısına göre sağlıklı ya da enfeksiyonlu olarak ikiye ayrılmışlar. Çalışmada laktasyon sırası, süt verimi, elektrik iletkenliği, sağım süresi ve kontrol mevsimi subklinik mastitisin belirlenmesinde kullanılan girdi değişkenleri olmuştur. Çalışmada, Holstein ineklerine ait bir veri kümesinde uygulanan destek vektör makinelerinin diğer hayvan popülasyonlarında da kolaylıkla uygulanabileceği vurgulanmıştır.

Ergülen ve Topuz (2008), çeşitli ırklarda süt verim tahminlerini yapay sinir ağları ile belirlemeyi amaçlamışlardır. Laktasyon süt verimine etki ettiği düşünülen, canlı ağırlık, cidago yüksekliği, ırk, yaş, günlük sağım sayısı ve mevsim özellikleri 90 hayvan üzerinde incelenmiştir. Tahmin edilen laktasyon süt değerleri ile gerçek ölçüm sonucu elde edilen laktasyon süt değerleri karşılaştırılmış, tahmin hatasının minimum olduğu görülmüştür. Ayrıca kullanılan 6 girişli tek çıkışlı ağ yapısının laktasyon süt veriminin tahmini için oldukça uygun bir yapı olduğu hesaplanan bulgularla ortaya konulmuştur.

Küçükönder ve ark. (2014) çalışmalarında Japon bildircini yumurtalarında döllülük üzerine etkisi olan mevsim, seleksiyon ve yerleşim sıklığı faktörlerinin veri madenciliği yöntemi ile sınıflandırılması ve bu faktörlerin etkisinin belirlenmesini amaçlamışlardır. Araştırmada kullanılan sınıflandırma algoritmaları sırasıyla yapay sinir ağları, radyal temel fonksiyonu sinir ağları, Bayes, KStar ve Ridor algoritmalarıdır. Yapılan bu çalışma ile % 99.73 doğru sınıflandırma başarısı ile bildircin yumurtalarının genel olarak % 85'inin döllu, % 15'nin ise üreme kapasitelerinin düşük olduğu tespit edilmiştir. Yumurtaların döllülük özelliğine göre (döllü-dölsüz) sınıflandırılması için veri madenciliğinde yer alan farklı algoritmalarından yararlanılmış ve yapılan karşılaştırma sonucunda Ridor algoritmasının en az hata ile daha başarılı sonuçlar ürettiği görülmüştür. Bu algoritmaya göre oluşturulan model ile yapılan sınıflandırma sonucunda, farklı mevsim ve yerleşim sıklıklarında

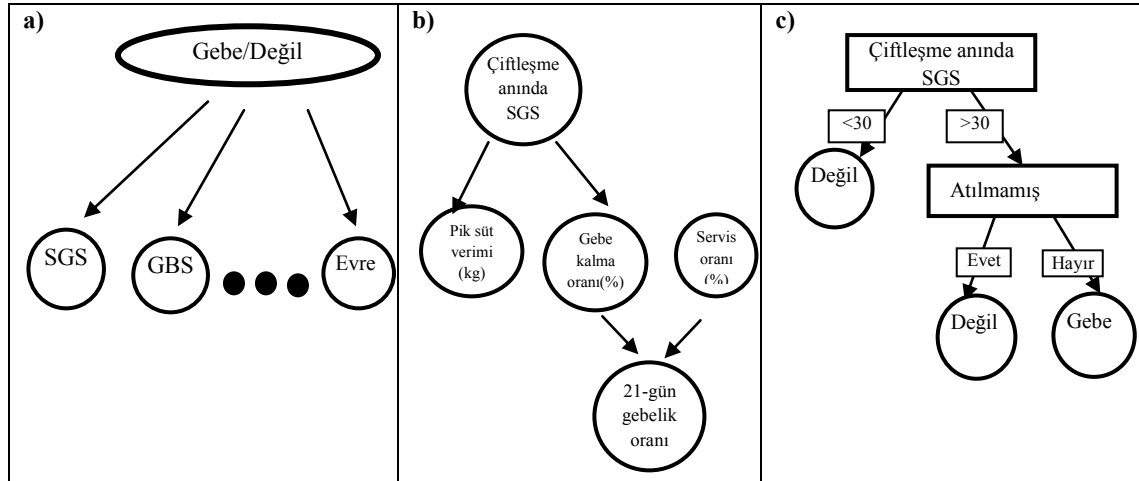
yetiştirilen iki hattan Japon bildircinlarına ait kuluçkalık yumurtalarda döllülük oranının % 85.71 olduğu belirlenmiştir.

Eyduran ve ark. (2013) yaptıkları çalışmada, farklı tarımsal veri kümelerine sınıflandırma ve regresyon ağacı metodlarını uygulamışlar ve sonuçların nasıl yorumlanacağını göstermişlerdir. Çalışmada, Pakistan’da yetiştirilen 138 Mengali kuzusu verisine sınıflandırma ve regresyon ağacı metodu uygulanmıştır. Veri değişkenleri olarak doğum ağırlığı, doğumun tipi (tek ve ikizlik), doğum yılı, cinsiyet, annenin yaşı, doğumda annenin ağırlığı kullanılmıştır. Sınıflandırma ağacı metodunda cinsiyet bağımlı değişken olarak seçilmiştir. Sonuç olarak doğum ağırlığı 3.75 kg’dan düşük olan tekli kuzular için erkek ve dişi olma oranları sırasıyla, % 28.6 ve % 71.4 olarak bulunmuştur. Regresyon ağacı metodu sonucunda yaşı 59 aydan büyük annelerin doğurduğu kuzuların en yüksek ortalama doğum ağırlığına (4 kg) sahip olduğu saptanmıştır. Çalışmanın sonucu olarak; sınıflandırma ve regresyon ağaçlarının çoklu bağlantı, aykırı değer, lineer olmayan problemler ve özellikle yüksek sayıda bağımsız değişken içeren karmaşık veri kümelerini değerlendirmede uygulanabilir yöntem olduğu belirtilmiştir.

Zidek ve ark. (2014) yaptıkları çalışmada iki farklı menşeli 416 baş Pinzgau sığırlarının DNA’larını kullanarak hayvanların kimlik teyidi amacı ile sınıflandırma modeli geliştirmişlerdir. Bunun için hayvanları menşelerine göre; 1. sınıfta

346, 2. sınıfta 66 hayvan olmak üzere 2 sınıfa ayırmışlardır. Hayvanların genetik yapıları analiz edilmiş ve 20 farklı veri madenciliği metodu ile modellenmiştir. 20 modelden 3 tanesi yüksek belirlilik değerlerine göre seçilmiştir. En küçük algoritma hatası karar ağaçları sınıflandırma tekniği uygulanan modelde elde edilmiştir (hayvanlar % 100 doğru sınıflandırılmıştır). Çalışmanın sonucu olarak veri madenciliğinde danişmanlı öğrenme araçlarının kullanımı ile genetik veriye dayanarak bilinmeyen örneklerin sınıflandırılabilme imkânı gösterilmiştir.

Shahinfar ve ark. (2014) yaptıkları araştırmalarında; 2000-2010 yılları arasında 300000’den fazla Holstein ineklerine ait üreme, üretim ve sağlık kayıtları kullanılarak tohumlama sonucunda gebe kalıp kalmama durumunu tahmin etmeye çalışmışlardır. Bu amaçla karar ağaçları, Bayes sınıflandırıcıları (NBC), torbalama (bagging), rastgele orman (random forest) ve Bayes ağları (bayesian networks) algoritmaları kullanılmıştır. Uygulanan 5 algoritma içinde rastgele orman diğerleri ile kıyaslamada önemli derecede iyi sınıflama yapmıştır. Şekil 8’de; Bayes sınıflandırıcıları, Bayes ağları ve Karar ağaçları algoritmalarının şeması gösterilmiştir. Araştırmacılar, laktasyonda olan bir süt sığırının tohumlanma sonucunun tahmin edilmesinin aşırı zor olmasına rağmen, hayvanın sağlığına, üreme geçmişine, üretim seviyesine ve çevresel özelliklere dair bilgilerin kullanılması ile yüksek üretkenliğe sahip hayvanların teşhisinin başarıyla yapılabileceğini belirtmişler.



Şekil 8. a) Bayes sınıflandırıcıları, b) Bayes ağları, c) Karar ağaçları algoritmalarının şeması (Shahinfar ve ark., 2014)

SGS: Sağımda geçen gün sayısı, GBS: Gönüllü bekleme süresi, Evre: Laktasyon evresidir.

4. Sonuçlar

Veri madenciliği, sayısal ve istatistiksel olarak büyük veri kümeleri üzerinde yoğun işlemler yapmayı gerektirir. Veri madenciliği çalışması esas olarak bir istatistik uygulamasıdır. Verilen bir örnek kümesine bir kestirici oturtmayı amaçlar. Bu çalışmada, veri madenciliğinin temel bilgileri verilmiş ve bu bilgiler çerçevesinde sınıflama ve kümeleme başlıkları altında çeşitli uygulamalara değinilmeye çalışılmıştır. Özellikle çok büyük hacimli veritabanlarında ya da birden fazla veritabanı içeren sistemlerde çalışırken bilgi keşif süreci çok önemlidir. Bu nedenle veri madenciliğiyle ilgili bir uygulamaya başlamadan önce veri madenciliğiyle ilgili temel kavramları ve kullanılan metotların iyi bilinmesi uygulamanın sağlıklı ve kısa sürede tamamlanmasını sağlayacaktır. Yapılan çalışma ile veri madenciliğinin özellikle süt ineklerinde süt verim tahminleri ve distosi belirleme çalışmaları başta olmak üzere, bildircin, broyler tavuklarda da kullanımına yer verilmiştir. Bu çalışmanın gelecekte yapılan çalışmalara bir temel ve fikir oluşturacağı öngörülmektedir.

Kaynaklar

- Abu-Hanna, A., De Keizer, N., 2003. Integrating classification trees with local logistic regression in Intensive Care prognosis. *Artificial Intelligence in Medicine*, 29(1-2): 5-23.
- Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York.
- Cortes, C., Vapnik, V., 1995. Support vector networks. *Machine Learning*, 20: 273-297.
- De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecological Society of America*, 81(11): 3178-3192.
- Elmas, Ç., 2011. Yapay Zeka Uygulamaları (2. Baskı). Seçkin Yayıncılık, Ankara.
- Ergülen, A., Topuz, D., 2008. İşletmelerdeki verimliliğin tahmin edilebilmesi ve bu verimliliği etkileyen faktörlerin MLP tipi yapay sinir ağıları tekniği ile belirlenmesi. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(8): 219-231.
- Eyduran, E., Tatlıyer, A., Tariq, M.M., Waheed, A., 2013. Application of classification and regression tree methods in agriculture. *Ulusal Tarım Kongresi*, 26-29 Ekim, Antalya.
- Görgülü, O., 2012. Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. *South African Journal of Animal Science*, 42(3): 280-287.
- Grzesiak, W., Lacroix, R., Wójcik, J., Blaszczyk, P., 2003. A comparison of neural network and multiple regression prediction for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science*, 83: 307-310.
- Grzesiak, W., Blaszczyk, P., Lacroix, R., 2006. Methods of predicting milk yield in dairy cows-Predictive capabilities of Wood's lactation curve and artificial neural networks (ANN). *Computers and Electronics in Agriculture*, 54(2): 69-83.
- Grzesiak, W., Zaborski, D., Sablik, P., Zukiewicz, A., Dybus, A., Szatkowska, I., 2010. Detection of cows with insemination problems using selected classification models. *Computers and Electronics in Agriculture*, 74(2): 265-273.
- Grzesiak, W., Zaborski, D., Sablik, P., Pilarczyk, R., 2011. Detection of difficult conceptions in dairy cows using selected data mining methods. *Animal Science Paper and Reports*, 29: 293-302.
- Grzesiak, W., Zaborski, D., 2012. Examples of the use of data mining methods in animal breeding. In: A. Karahoca (Ed), *Data mining applications in engineering and medicine*, InTech.
- Hartigan, J., 1975. Clustering Algorithms. John Wiley&Sons, New York.
- Kahramanlı, H., 2008. Hibrit bulanık sinir ağını kullanarak bir sınıflandırma ve kural çıkartma sisteminin geliştirilmesi. Doktora tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Kamphuis, C., Mollenhorst, H., Feelders, A., Pietersma, D., Hogeveen, H., 2010. Decision-tree induction to detect clinical mastitis with automatic milking. *Computers and Electronics in Agriculture*, 70(1): 60-68.
- Küçükönder, H., Üçkardeş, F., Nariç, D., 2014. Hayvancılık alanında bir veri madenciliği uygulaması: Japon bildircini yumurtalarında döllülüğe etki eden bazı faktörlerin belirlenmesi. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 20(6): 903-908.
- Larose, D.T., 2006. Data Mining Methods and Models. John Wiley&Sons, Inc., Hoboken, New Jersey.
- Lee, T.S., Chiu, C.C., Chou, Y.C., Lu, C.J., 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data Analysis*, 50: 1113-1130.
- Maldonado-Castillo, I., Eramian, M.G., Pierson, R.A., Singh, J., Adams, G.P., 2007. Classification of bovine reproductive cycle phase using ultrasound-detected features. *Fourth Canadian Conference on Computer and Robot Vision*, May 28-30, Montreal, Quebec, Canada, pp. 258-265.
- Mammadova, N., 2012. Süt sığırlarında mastitisin bazı yapay zeka yöntemleri kullanılarak erken dönemde tespiti. Doktora tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Mammadova, N., Keskin, İ., 2013. Application of the support vector machine to predict subclinical mastitis in dairy cattle. *The Scientific World Journal*, 1-9.
- Mucherino, A., Papajorgji, P., Pardalos, P.M., 2009. A survey of data mining techniques applied to agriculture. *International Journal of Operational Research*, 9(2): 121-140.

- Nispert, R., Elder, J., Miner, G., 2009. Statistical Analysis and Data Mining Application. Elsevier, USA.
- Piwczyński, D., 2009. Using classification trees in statistical analysis of discrete sheep reproduction traits. *Journal of Central European Agriculture*, 10: 303-310.
- Rajesh, D., 2011. Applied of spatial data mining for agriculture. *International Journal of Computer Applications*, 15(2): 7-9.
- Shahinfar, S., Page, D., Guenther, J., Cabrera, V., Fricke, P., Weigel, K., 2014. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *Journal of Dairy Science*, 97(2): 731-742.
- Takma, Ç., Atlı, H., Aksakal, V., 2012. Çoklu doğrusal regresyon ve yapay sinir ağı modellerinin laktasyon süt verimine uyum yeteneklerinin karşılaştırılması. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 18(6): 941-944.
- Wood, P.D.P., 1967. Algebraic model of the lactation curve in cattle. *Nature*, 216: 164-165.
- Vale, M.M., Moura, D.J., Naas, I.A., Oliveira, S.R.M., Rodrigues, L.H.A., 2008. Data mining to estimate broiler mortality when exposed to heat wave. *Scientific Agriculture*, 65(3): 223-229.
- Zaborski, D., Grzesiak, W., 2011a. Detection of heifers with dystosia using artificial neural networks with regard to ER α -BgII, ER α -SnaBI and CYP19-PvuII genotypes. *Acta Scientiarum Polonorum Zootechnica*, 10(2): 105-116.
- Zaborski, D., Grzesiak, W., 2011b. Detection of difficult calvings in dairy cows using neural classifier. *Archiv Tierzucht*, 54(5): 477-489.
- Zareipour, H., Bhattacharya, K., Canizares, C.A., 2006. Forecasting the hourly Ontario energy price by multivariate adaptive regression splines. IEEE, Power Engineering Society General Meeting, pp. 1-7.
- Židek, R., Šidlová, V., Kasarda, R., Fuerst, W., 2014. Methods for distinction of cattle using supervised learning. *International Scholarly and Scientific Research Innovation*, 8(5): 498-500.