



**Volkan TÜRKMEN<sup>1a\*</sup>**  
**Doğukan ÖZEN<sup>2b</sup>**

<sup>1</sup>Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Ankara

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

**ORCID<sup>a</sup>:** 0000-0003-4215-4203  
**ORCID<sup>b</sup>:** 0000-0003-1943-2690

**\*Sorumlu Yazar:** Volkan TÜRKMEN  
**E-Posta:** volkantrkmen@yandex.com

**Geliş Tarihi:** 03.06.2023  
**Kabul Tarihi:** 30.11.2023

**14 (3): 118-132, 2023**  
**DOI: 10.38137/vftd.1307581**

#### **Makale atıf**

*Türkmen, V. ve Özen, D. (2023). Makine Öğrenmesinde Regresyon Algoritmalarının Veteriner Hekimliği Alanında Uygulamaları, Veteriner Farmakoloji ve Toksikoloji Derneği Bülteni, 14 (3), 118-132. DOI: 10.38137/vftd.1307581.*

## **MAKİNE ÖĞRENMESİNDE REGRESYON ALGORİTMALARININ VETERİNER HEKİMLİĞİ ALANINDA UYGULAMALARI**

**ÖZET.** Hızla artan veriler, özellikle giyilebilir teknolojiler, sensörler ve internet bağlantılı akıllı ürünler (IoT) gibi yeni veri kaynaklarından akan daha büyük, daha karmaşık veri kümeleri makine öğrenmesi algoritmalarına olan ihtiyacı her geçen gün artırmaktadır. Geleneksel veri işleme yazılımlarının başa çıkamadığı büyük hacimli veri kümeleri her alanda olduğu gibi veteriner hekimlik alanında da yeni fırsatlar sunmakta ve daha önce üstesinden gelinemeyen sorunlar için yeni çözüm yolları üretebilme potansiyeline sahip olduğu görülmektedir. Bu derleme çalışmasında literatürde sıkça karşılaşılan makine öğrenmesi regresyon algoritmaları tanıtılmış ve veteriner hekimliği alanında uygulamalarına ilişkin kısa örnekler verilmiştir. Bunun yanında bu algoritmaların veteriner hekimliği alanındaki potansiyeline kısaca değinilmiştir.

**Anahtar Kelimeler:** Makine öğrenmesi, regresyon algoritmaları, veteriner hekimlik.

## **APPLICATIONS of REGRESSION ALGORITHMS in MACHINE LEARNING in VETERINARY MEDICINE**

**ABSTRACT.** Machine learning algorithms are becoming more and more essential as a result of the exponential growth of data, particularly the larger, more complicated datasets coming from emerging data sources like wearables, sensors, and internet-connected smart devices (IoT). In the field of veterinary medicine, as in every discipline, large volumes of data sets that conventional data processing tools cannot handle present new opportunities and have the ability to generate novel solutions for issues that have previously been insurmountable. Machine learning regression methods, which are widely found in the literature, are described in this study along with a few brief examples of how they might be used in veterinary medicine. Also, a brief mention is made of how these algorithms could have applications in veterinary medicine.

**Keywords:** Machine learning, regression algorithms, veterinary medicine.

## GİRİŞ

Veteriner hekimliği; hayvan sağlığı, halk sağlığı, zoonotik hastalıklar gibi konuları içeren oldukça geniş ve büyüyen bir disiplindir. Benzer şekilde yapay zeka (AI, Artificial Intelligence.); felsefe, matematik, sinir bilimi, kontrol teorisi ve sibernetik, bilgisayar mühendisliği ve veri bilimleri dahil olmak üzere birçok bilim alanı ile ilişkilidir. Bu iki geniş ve büyüyen alanın kesişimi, birinin diğerini çarpıcı biçimde etkileme potansiyeline sahiptir. Dolayısıyla yapay zekanın veteriner hekimliği alanındaki uygulamaları neredeyse sınırsızdır. Benzer şekilde veteriner hekimliği alanındaki yapay zeka uygulamaları da diğer alanlardaki yapay zeka ilerlemelerine etki edebilme potansiyeline sahiptir (Basran ve Appleby, 2022).

Yapay zeka çalışmaları, konuşma tanımlama (speech recognition), bilgisayarla görme (computer vision) ve doğal dil işleme (natural language processing) olmak üzere genellikle üç büyük kategoriye ayrılırlar. Veteriner hekimlikte bu amaçlardan herhangi biri için yapay zekadan nasıl yararlanılabileceği, belirli göreve ve mevcut veri türlerine bağlıdır. Her çalışma alanı, veterinerlik tıbbında AI için sayısız uygulama sunma potansiyeline sahiptir (Basran ve Appleby, 2022).

Hayvan sağlığı alanında tanı, teşhis ve tedavide veteriner hekimlerin hızlı çözümler üretebilmesi veya doğru kararlar alabilmesi için yapay zekanın sunduğu teknolojik imkanlardan faydalanması gerekliliği görülmektedir. Zettabayt boyutlarında ciddi miktarlardaki verinin akmaya başladığı günümüzde büyük boyutlu veri setlerinden makine öğrenmesi algoritmaları ile öğrenilen desenler ve modeller ile küçük yapay zekalar oluşturup örneğin hayvan yetiştiriciliğinde verimi, maliyeti, kaliteyi, hastalıkları sıkı bir şekilde takip etmemizi sağlayabilir. Her koşulda veteriner hekimlerin daha hızlı karar almasına yardımcı olabilir (Cihan ve ark., 2017).

Yapay zekanın alt dallarından biri olan makine öğrenmesi (ML, Machine Learning), matematiksel ve istatistiksel yöntemler kullanarak mevcut verilerden çıkarımlar yapan, bu çıkarımlarla bilinmeyene dair tahminlerde bulunan modelleme ve algoritmalarından oluşmaktadır (Kıral Özkan, 2015). Bir başka ifade ile

ML, genellikle geleneksel veri işleme araçları ile analizi yapılamayan ve karmaşık bir yapıya sahip büyük miktardaki veri setlerini kullanarak anlamlı ilişkiler bulan model ve desenler geliştiren yöntemler ailesidir (Chen ve He, 2014). ML denetimli, yarı denetimli, denetimsiz ve takviyeli algoritmalar gibi yöntemleri kullanarak, belirli bir görev ve bununla ilişkili model desenlerini ortaya çıkarabilirler.

Denetimli ML algoritmaları, tüm veriler daha önce sınıflandırıldığında kullanılır. Burada amaç, yeni girdi verilerini bilinen çıktı yanıtlarına göre sınıflandırmaktır. Çıktılar kategorik, sıralı kategorik veya sürekli olabilir. Denetimsiz öğrenme algoritmaları ise, öncelikle veri çıkarımı için veya veri kümesinden makul sonuçlar veya potansiyel hipotezler çıkarmak için kullanılmaktadır. Burada amaç genellikle büyük veri kümelerini, anlamlı ve yorumlanabilir halde gruplamaktır. Denetimli ML'de olduğu gibi, veri türleri kesikli veya sürekli olabilir. Diğer ML örnekleri arasında yarı denetimli ML bulunur. Yarı denetimli ML, hem denetimli hem de denetimsiz ML'yi birleştiren ve etiketli ve etiketsiz veri kümelerinin bir kombinasyonu olduğunda kullanışlı olan yapay zeka dalıdır. Takviyeli (pekiştirmeli) öğrenme yöntemini kullanan algoritmalar öğrenme performansını artırmak için çevre ile etkileşim kuran bir sisteme dayanmaktadır. Bu öğrenme yöntemi veri setindeki mevcut bilgiyi kullanma ve deneme yanılma yoluyla keşfetme arasında bir denge kurmaya çalışır. Canlıların öğrenme yöntemlerini taklit eden bir yol izleyerek algoritmaların öğrenmesini sağlar. Pekiştirmeli öğrenmenin denetimli öğrenmeden ayrıldığı noktalar şunlardır; doğru etiket veya değer eşleşmesi verilmemekte, uygun olmayan eylemler için dışarıdan düzeltme yapılamamaktadır (Kıral Özkan, 2015).

Bu derleme çalışmasının temel amacı makine öğrenmesinde kullanılan regresyon algoritmalarının tanıtılması ve veteriner hekimliği alanında uygulamalarına ilişkin kısa örnekler ile bu algoritmaların veteriner hekimliği alanındaki potansiyelini göstermektir. Bu amaçlar doğrultusunda ilerleyen başlıklarda ilk etapta kısaca makine öğrenmesinin analiz kısmında yapılacak işlemler, veri ön işleme, öznelik seçimi, hiper parametre en iyilemesi, model performans değerlendirme yöntemleri ve ölçütler açıklanmıştır. Sonrasında ise gerçekleştirilen literatür taraması ile veteriner hekimliğinde makine öğrenimi regresyon algoritmaları kullanılan çalışma örnekleri kısaca izah edilerek çalışma bitirilmiştir.

## REGRESYONA DAYALI DENETİMLİ MAKİNE ÖĞRENMESİ ALGORİTMALARI

Regresyon çözümlenmesi genel olarak bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişki yapısını en iyi matematiksel modeller ile açıklamaya çalışan yöntemler ailesidir. Makine öğrenmesi yapay zekayı oluşturan en önemli bileşenlerden birisidir. İstatistiksel öğrenme üzerine kurulu makine öğrenmesi, bir sistemin çevresine uyum sağlayabilmek için kendi kendini yeniden düzenlemesi olarak nitelendirilen öğrenme işinin bilgisayarlar tarafından gerçekleştirilmesini sağlamak için kullanılır (Kıral Özkan, 2015). Regresyon algoritmaları denetimli öğrenme yöntemi ile makinelerin eğitildiği makine öğrenmesi algoritmalarıdır.

Çalışmanın bu bölümünde literatürde sıklıkla kullanılan klasik makine öğrenmesi regresyon yöntemleri olan doğrusal regresyon, Ridge, LASSO (Least Absolute Shrinkage and Selection Operator) ve Elastik Net (Elastic Net) gibi düzenleme algoritmaları ile hem sınıflandırma hem de regresyon problemleri için kullanılma imkanı bulunan yapay sinir ağları (YSA), Naïve Bayes, karar ağaçları, k-en yakın komşu (KNN, k-nearest neighbors), destek vektör makineleri (DVM) ve destek vektör regresyonu (DVR) algoritmaları tarihsel gelişim sürecine göre sırasıyla açıklanacaktır. Sonraki aşamada ise günümüzde yaygın olarak kullanılan topluluk öğrenme yöntemleri; Torbalama (Bagging), Güçlendirme-Hızlandırma (Boosting), Gradyan Arttırıcı (Gradient Boosting) ve Rastgele Orman (Random Forest) tanıtılmıştır (Ceylan, 2018).

## DOĞRUSAL REGRESYON

Basit ve çok değişkenli doğrusal regresyonun temel amacı, bağımlı ve bağımsız değişken(ler) arasındaki doğrusal ilişkiyi matematiksel modellerle açıklayan bağıntı(ları) bulmak ve bu model yardımıyla bağımlı değişken hakkında ileriye yönelik tahminlerde bulunmaktır. Regresyon çözümlenmesi geleneksel olarak istatistik paket programlarında en küçük kareler (EKK) yöntemi kullanılarak hata kareler toplamını en aza indirerek yapılır. Basit doğrusal regresyon, kullanışlı ve çok geniş kullanım alanı olan bir istatistiksel öğrenme metodudur (James ve ark., 2013).

Doğrusal regresyonun makine öğrenmesi algoritması olarak uygulamasında meyilli azalım (Gradient Descent-GD) yöntemi kullanılır. Makine öğrenmesinde hata oranını en aza indirmek için sıklıkla kullanılan en iyileme (optimizasyon) yöntemidir. Bu yöntem, rastgele alınan değişkenlerle başlayarak global minimum değerine ulaşmayı amaçlayan iteratif bir yöntemdir. Öğrenme oranının (learning rate) eklendiği kayıp fonksiyonu kullanılarak hatayı veya kaybı en aza indirmek için regresyon katsayıları güncellenir, bu süreç global minimum değerine ulaşmaya kadar devam eder. Farklı regresyon türleri için farklı kayıp fonksiyonları kullanılır. Doğrusal regresyon için ortalama hata kareler (MSE, Mean Squared Error) kayıp ya da maliyet fonksiyonu olarak kullanılır (Gandhi, 2018).

## DÜZENLİLEŞTİRME (REGULARIZATION) ALGORİTMALARI

Çok değişkenli doğrusal regresyon analizi sonucunda varyans artırıcı faktör (VIF, Variance Inflation Factor) 10'un üstünde ve koşul indeksi 30'un üstünde ise bağımsız değişkenler arasında kuvvetli bir ilişkinin varlığından söz edilebilir. Çoklu bağlantı sorunu olarak bilinen bu durum sebebi ile regresyon modeli hatalı tahminler üretir, katsayılar gerçekte anlamlı iken anlamsız çıkabilir. Bu durumun çözümü için literatürde en çok tercih edilen ve yanlış tahmin regresyonları olarak da bilinen Ridge, LASSO ve Elastik Net düzenleme algoritmaları kullanılır (Alpar, 2011).

Bu üç teknik karmaşıklıklarına göre modelleri cezalandıran ve aşırı uyum (over-fitting) sorununu önlemek için sıklıkla kullanılan basit düzenleme makine öğrenmesi algoritmalarıdır. Ridge regresyon Arthur Hoerl ve Robert Kennard (1970) tarafından önerilen L2 (Tikhonov düzenlemesi) türü bir düzenleme algoritmasıdır. Ridge regresyon yönteminde maliyet fonksiyonuna (cost function) düzenleme terimi olarak katsayıların karesi eklenir;

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}.$$

Tibshirani tarafından 1996 yılında önerilen LASSO regresyon yönteminde maliyet fonksiyonuna düzenleme terimi olarak katsayıların mutlak değeri eklenir;

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}.$$

Elastik Net regresyon yöntemi Zou ve Hastie (2005) tarafından LASSO yönteminin bazı eksikliklerine çözüm bulmak için önerilmiştir. Elastik Net yönteminde maliyet fonksiyonuna düzenleme terimi olarak ise her ikisinin toplamı eklenir;

$$\lambda \sum_{j=1}^p (\alpha \beta_j^2 + (1 - \alpha) |\beta_j|)$$

ve sonra büzülme (shrinkage) ve cezalandırma (penalize) terimi olarak öğrenme oranı eklenerek model yeniden düzenlenir (Jason, 2019).

Kayıp fonksiyonundaki hata oranını en aza indirerek global minimum değerine ulaşmak için iteratif bir metot olan GD en iyileme (optimizasyon) yöntemi kullanılır.

### YAPAY SİNİR AĞLARI ALGORİTMASI

Yapay Sinir Ağları (YSA) Warren McCulloch ve Walter Pitts (1943) tarafından geliştirilmiştir. YSA insan beynindeki nöronlardan esinlenerek oluşturulmuştur. İnsan beyninin çalışma prensiplerinden faydalanılarak nöron isimli sinir hücrelerinin davranışları modellenerek geliştirilmiştir. Birbirlerine çoklu bağlantı yapmış nöronların oluşturduğu karmaşık bir yapı halinde paralel işlemciler gibi çalışan çok yönlü bir makine öğrenmesi tekniğidir (Balta, 2018). Sınıflandırma, regresyon, denetimsiz öğrenme, örüntü tanıma vb. alanlarda kullanılabilir.

YSA' yı oluşturan temel unsurlar şu şekildedir:

1. Bilgi işlemenin gerçekleştiği nöron adı verilen birçok temel birimden oluşur (mimari yapı).
2. Nöronlar arası bağlantı hatları boyunca sinyaller iletilir (algoritma, eğitme veya öğrenme).
3. Her bir bağlantı hattı bağlılık durumunu gösteren sayısal bir ağırlığa sahiptir. Bu ağırlık sinyal değeri ile çarpılır.
4. Her bir nöron, giriş sinyallerinin toplamını genellikle doğrusal olmayan bir aktivasyon fonksiyonuna giriş olarak kullanıp bir çıkış sinyali üretir (aktivasyon fonksiyonu).

Çok katmanlı ağlar, genellikle üç katmanlı bir yapıda veya daha fazla katman kullanılarak oluşturulan YSA modellemesidir. Giriş katmanı gerçek değerleri

giriş olarak alan nöronlardan oluşur. Gizli katman ise giriş katmanından gelen sinyaller ile beslenen nöronlardan oluşur. Çok katmanlı YSA' da kullanılan çeşitli aktivasyon fonksiyonları vardır. Bunlardan en yaygın olanı Sigmoid olarak adlandırılan fonksiyondur (Balta, 2018). Aktivasyon fonksiyonlarındaki amaç; ağırlık ve eşik değerlerini ayarlamaktır. Sigmoid, tanh, relu, leaky relu ve elu gibi birçok aktivasyon fonksiyonu bulunmaktadır. Günümüzün en popüler aktivasyon fonksiyonu relu fonksiyonudur.

### NAİVE BAYES ALGORİTMASI

18.yy'da olasılık üzerine çalışan Thomas Bayes'in tarafından geliştirilen Bayes teoremine dayanan algoritmadır. Naive Bayes algoritması 1961 yılında tanıtılmıştır (Maron, 1961). Naive Bayes geçmiş olaylara ait verileri kullanarak gelecekteki olayların olasılıklarını tahmin etmeyi amaçlayan yöntemdir. Naive Bayes algoritmasının, "Naive (naif, saf)" şeklinde isimlendirilmesinin nedeni, algoritmanın veri hakkında yapmış olduğu bir takım ön kabullerden kaynaklanmaktadır. Söz konusu algoritma, veri setindeki tüm özelliklerin eşit derecede önemli olduğunu ve özelliklerin birbirinden bağımsız olduğunu farz etmektedir (Bulut, 2019).

Bu algoritma frekans tabloları vasıtasıyla veri setini öğrenmekte, her bir özelliğin kategorik olması sayesinde sınıflar ve özellikler arasında kombinasyonlar yapılarak matris oluşturulabilmektedir. Ancak nümerik özellikler kategorik değerlere sahip olmadığından, algoritmanın sayısal verilerle doğrudan çalışması mümkün olmamaktadır. Bu sorunu aşmak için nümerik özellikler kategorik kümelere dönüştürülür. Verilerin dağılımına göre kesim noktaları belirlenir, böylelikle sayısal veriler için kategoriler oluşturulur (Bulut, 2019).

Naive Bayes algoritması, basit, hızlı ve efektiftir. Anlamsız ve kayıp verilerin bulunduğu veri setlerinde başarılıdır. Göreceli olarak eğitim aşamasında az sayıda örnek veriye ihtiyaç duyar. Öngörü yapılması için gerekli olan tahmini olasılığın elde edilmesi kolaydır. Özelliklerin birbirinden bağımsız olduğu ve eşit derecede önemli olduğu gibi genellikle yanlış bir ön kabul yapmaktadır. Sayısal verilerin fazla olduğu veri setleri için ideal bir yöntem değildir (Bulut, 2019).

### KARAR AĞAÇLARI ALGORİTMASI

Ağaç tabanlı öğrenme algoritmaları sınıflandırma ve regresyon problemlerinin çözümünde en çok kullanılan

denetimli öğrenme yöntemlerinden biridir. Doğrusal modellerin yanında, doğrusal olmayan ilişkileri de oldukça iyi gösterir. Parametrik olmayan örnek tabanlı (instance-based) makine öğrenmesi algoritmasıdır. Regresyon karar ağaçları Morgan ve Sonquist (1963) tarafından geliştirilmiştir.

Ağaç, dallardan ve düğümlerden oluşur. Karar Ağacı bölünmeye tüm gözlemleri barındıran kök (root) düğümden başlar. Yukarıdan aşağıya tekniği ile genellikle açgözlü veya iştahlı (greedy) yaklaşımla yapılır. Bu yöntem ile her aşamada opsiyonel bazı kriterlere göre ağaç dallara ayrılır. Bu yaklaşımla aşağıya doğru dallara ayırma yaparken en iyi bölünmeyi sağlayacak değişkeni seçmek için “Entropi” ve “Bilgi kazancı” ile “Gini indeksi” ve “Karışıklık ölçüsü (measure of impurity)” kriterleri kullanılmaktadır. Bilgi kazancı en yüksek değişken en iyi dallara ayırmayı sağlayacak değişken olarak seçilir ve bölünmeye o değişkenden başlanılır. Bu bölünme, en çok ağaç derinliği ve bir düğümden bölünme için ele alınan en küçük eleman sayısı gibi belirli durdurma kuralına kadar devam eder (Akman, 2010).

Karar ağaçları kurulma biçimlerine göre farklılıklar gösterirler. Kullanılan algoritmaya (ID3, C4.5, CART vb.) göre oluşturulan ağacın şekli değişebilir. Değişik ağaç yapıları da farklı sınıflandırma sonuçları verir (Yılmaz, 2014).

Karar ağacı regresyonu bağımsız değişkenleri bilgi kazancına göre aralıklara ayırır. Karar ağacı regresyonu diğer regresyon modellerinde olduğu gibi sürekli değil, kesikli sonuç üretir. Eğitim sırasında elde edilen regresyon modeli kesikli sonuç ürettiği için model test edilirken belirli bir aralıktaki değer tahmin edilmek istenildiğinde bu aralıktaki ortalamayı sonuç olarak üretir. Yani bu aralıktaki sonuçlar aynı sonucu (ortalamayı) verir (Şirin, 2017).

Oluşturulan karar ağacına budama (pruning) işlemi uygulanır. Budama işlemi, aşırı uyum problemini engellemek ve yanlış sınıflandırma hatasını en küçük yapmak için ağacın gerçek büyüklüğüne karar vermede kullanılır (Özgür, 2020).

### **K EN YAKIN KOMŞU (KNN, K NEAREST NEIGHBOUR) ALGORİTMASI**

KNN algoritması, 1967 yılında T. M. Cover ve P. E.

Hart tarafından önerilmiştir. Uygulaması basit ve esnek bir sınıflandırma ve regresyon yöntemidir. Genellikle öğrenme verilerinin sayısal olduğu durumlarda kullanılır. Öğrenmekten ziyade veri kümesini bütünüyle hafızaya alıp işlem yapar. Veriler hakkında herhangi bir parametre öğrenilmediğinden parametrik olmayan bellek tabanlı bir makine öğrenmesi yaklaşımı olarak bilinir. Bir model inşa etmediği için literatürde bu tarz öğrenme şekline tembel öğrenme "lazy learning" de denir. Kategorik özelliklerin olduğu durumlarda bu veriler, uygun şekilde kodlanıp sayısal hale getirilerek sınıflandırma işlemi yapılabilir (Balta, 2018). Bağımlı ve bağımsız değişken arasındaki ilişkinin doğrusal olmadığı durumlarda doğrusal regresyon analizinden daha iyi sonuçlar vermektedir (James ve ark., 2013).

KNN algoritmasında k (komşu noktalar) parametresinin seçimi, genelleştirme problemi açısından önemli bir yer tutmaktadır. K büyük seçilirse gürültülü verilerin etkisi azalır, küçük seçilirse anlamsız ve aykırı örneklerin model üzerindeki etkisi artar (Bulut, 2019).

KNN algoritması, k tane benzer örneği seçerken bir uzaklık (distance) fonksiyonu kullanmakta, bu fonksiyon vasıtasıyla iki örnek arasındaki uzaklığı hesaplamaktadır. Uzaklığı hesaplamak için farklı yaklaşımlar bulunmakla birlikte, geleneksel olarak kNN, Öklid mesafesini kullanmaktadır (Bulut, 2019).

### **DESTEK VEKTÖR MAKİNELERİ**

Destek Vektör Makinelerinin (DVM), teorik temelleri V. Vapnik tarafından 1960'lı yıllarda ortaya atılmıştır. DVM algoritması Boser, Guyo ve Vapnik tarafından 1992 yılında yayınlanan çalışma ile tanıtılmıştır. Sınıflandırma ve regresyon için uygulanabilir denetimli öğrenme tekniklerindedir. Bu yöntemin temeli istatistiksel öğrenme teorisine ve yapısal risk minimizasyonuna dayanmaktadır (Ceylan, 2018).

DVM, temel olarak iki sınıfı bir doğru veya düzlem ile birbirinden ayırmaya çalışır. Bu ayırmayı da sınırdaki elemanlara göre yapar. Sınıfların doğrusal ayrılabilirdiği ve ayrılamadığı durumlarda kullanılabilme, model karmaşıklığını ayarlayabilme ve genelleme yeteneği gibi özellikleri sayesinde birçok alanda kullanımı tercih edilen bir yöntem olmuştur. Sınıfların doğrusal ayrılamadığı durumlarda karar sınırlarını oluşturmak için çekirdek hilesi (kernel trick) kullanır (Ceylan, 2018).

DVM'ler farklı sınıf etiketlerini birbirinden ayıran çok

boyutlu alanda hiper düzlemler oluşturarak sınıflandırma problemlerinde kullanılmak üzere geliştirilen bir metottur. Optimal bir hiper düzlem oluşturmak için hata fonksiyonunu en aza indirmek ve marjini maksimuma çıkartmak amacı ile aşamalı bir eğitim algoritması kullanır (Ceylan, 2018).

DVM'ler sınıflandırma problemlerini çözmenin yanı sıra regresyon problemlerine de uygulanabilir. Destek vektör regresyonu (DVR), regresyon tahminleri için kullanılan bir tür destek vektör makinesidir. DVR'nin amacı veri noktalarının olabildiği kadar yakın olduğu hiper düzlem oluşturmaktır (Trafalis ve İnce, 2000). Sürekli değişken olan çıktı (output) değerini en iyi tahmin eden fonksiyonu bulmaya çalışır. En küçük kareler, Laplace, Huber ve  $\varepsilon$ -insensitive alan yazında en çok tercih edilen kayıp fonksiyonlarıdır. Kayıp (loss) fonksiyonu bir mesafe ölçüsü içerecek şekilde değiştirilir. DVR için en çok tercih edilen  $\varepsilon$ -insensitive ( $L_\varepsilon$ ) kayıp fonksiyonu şu şekilde gösterilir (Smola ve Scholkopf, 2004);

$$L_\varepsilon(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon & \text{if } |f(x) - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

$\varepsilon$  pozitif bir sayıdır ve  $f(x)$  fonksiyonu etrafındaki mesafeyi simgeler.  $Y$  gerçek bir değerdir. Eğer  $f(x)$  ile  $y$  arasında mesafe  $\varepsilon$  ile belirtilen pozitif bir sayıyı geçmiyorsa  $f(x)$  ile  $y$  arasında bir kayıp yoktur. DVR algoritması doğrusal regresyon problemi şeklinde formüle edildiğinde aşağıdaki şekilde yazılır (Smola ve Scholkopf, 2004);

$$f(x) = \langle w, x \rangle + b \text{ with } w \in \mathcal{X}, b \in \mathbb{R}$$

$w$  ağırlık vektörünü,  $\langle \cdot, \cdot \rangle$  nokta çarpımı (dot product) gösterim şeklini ve  $b$  sabit bir sayıyı belirtir.  $\varepsilon$ -insensitive ( $L_\varepsilon$ ) kayıp fonksiyonu formülasyonu kullanılarak aşağıdaki düzenlenilmiş risk fonksiyonu elde edilir. Bu fonksiyon en küçüklenecek  $w$  ağırlık vektörü ve  $b$  sabit sayısı hesaplanabilir (Smola ve Scholkopf, 2004).

$$\begin{aligned} &\text{minimize} && \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ &\text{subject to} && \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned}$$

Vapnik (1995) tarafından belirtilen yukardaki formülasyonda  $\frac{1}{2} \|w\|^2$  ( $\|w\|^2 = \langle w, w \rangle$ ) güven aralığını,  $\xi_i, \xi_i^*$  gevşek (slack) değişkeni ve  $C$  düzenleme sabitini belirtir. Problemi çözmek için Lagrange çarpanı ve KKT (Karush-Kuhn-Tucker) durumu kullanılarak aşağıdaki Lagrange ikilemi (dual Lagrangian) problemi formülasyonu elde edilir (Wang ve ark. 2012).

$$\begin{aligned} &\min && \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j^* - \alpha_j) K(x_i \cdot x_j) + \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ &\text{s.t.} && \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ &&& 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, l \end{aligned}$$

Yukardaki problem çözüldüğünde  $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)^T$  ve  $w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i$  için en iyi (optimal) ve ideal sonuç elde edilir. Doğrusal regresyon formu aşağıdaki şekilde yazılabilir (Wang ve ark. 2012).

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \bar{\alpha}_i) K(x_i \cdot x) + \bar{b}$$

$K(x_i \cdot x)$  kernel fonksiyonunu gösterir. Bu formülasyon destek vektör açılımı (Support Vector expansion) olarak adlandırılır (Wang ve ark., 2012).

DVR hem doğrusal hem de doğrusal olmayan çekirdek fonksiyonlarını kullanabilir. Doğrusal bir çekirdek, iki giriş vektörü arasındaki basit bir nokta ürünüdür. Doğrusal olmayan çekirdek fonksiyonları ise verilerdeki daha karmaşık modelleri ortaya çıkarabilen karmaşık fonksiyonlardır. Çekirdekler görevin karmaşıklığına ve verinin özelliklerine göre seçilir (Salunke, 2023).

DVM'ler doğrusal olmayan sınıflandırma veya regresyon problemlerinde maksimum marjlı hiper düzlemi bulmak için elimizdeki koordinatları belirli çekirdek fonksiyonları ile çarparak çok daha anlamlı hale getirebiliriz. Düşük boyutlar karmaşık veri setlerini açıklamada yeterli olmayabilir. Çekirdek fonksiyonları yardımıyla girdiler daha yüksek boyutlu bir uzaya taşınır ve burada doğrusal olarak ayrılabilir hale gelir (Ceylan, 2018).

Çekirdek fonksiyonları çeşitli avantajlar sağlar. Veri dönüşümüne ihtiyaç duymadan doğrusal olmayan sınıflandırma veya regresyon görevlerini yerine getirebilirler. Bunun yanında, verilerin belirli özelliklerine göre seçilebilen ve çok çeşitli doğrusal olmayan dönüşüm türleri sağlayabilen

farklı çekirdek fonksiyonlarının kullanılmasına izin verirler (Salunke, 2023).

En yaygın kullanılan çekirdek fonksiyonları; doğrusal, polinom, Gauss, üstel (Exponential), radyal tabanlı fonksiyon (RBF, Radial Basis Function) ve Laplace RBF'dir. Doğrusal çekirdek doğrusal sınıflandırma veya regresyon tahminleri için, polinom çekirdeği doğrusal olmayan problemler için ve RBF çekirdeği genellikle çok sayıda özneliğe sahip sınıflandırma problemleri için kullanılır (Salunke, 2023).

### TOPLULUK (ENSEMBLE) ÖĞRENME ALGORİTMALARI

Topluluk öğrenme yöntemlerinde topluluk, çeşitli modeller oluşturularak ve tahminler birleştirilerek oluşturulur. Farklı birleştirme yöntemleri ile bir araya getirilen zayıf öğrencilerin bir araya getirilmesi ile oluşturulan güçlü öğrencilerin oluşturduğu topluluğun öğrenme yöntemidir. Tahmin ve sınıflandırma doğruluğunu arttırmak için karar ağacı gibi tek karar vericilerin bir araya getirilmesi ile oluşur. Tahmin varyansını ve yanlılığını azaltmak için tercih edilen algoritmaları içinde barındıran öğrenme yöntemidir. Kaggle gibi internet ortamında yapılan yarışmalarda problemi çözme hızından ve doğruluk oranının yüksekliğinden dolayı en çok tercih edilen algoritmaları içinde barındırır. Birden çok modelin tahminleri kullanıldığından algoritmaların doğruluk oranı daha yüksek ve başarılıdır. En yaygın olarak kullanılan topluluk öğrenme yöntemlerine Torbalama (Bagging), Güçlendirme-Hızlandırma (Boosting), Rastgele Orman (Random Forest) ve Gradyan Arttırıcı (Gradient Boosting) örnek verilebilir (Ceylan, 2018).

### GÜÇLENDİRME-HIZLANDIRMA ALGORİTMASI (BOOSTING)

Güçlendirme-hızlandırma veya yükseltme olarak adlandırılan algoritma Robert Schapire ve Yoav Freund tarafından geliştirilmiştir. İlk kanıtlanabilir ve etkili güçlendirme-hızlandırma algoritması Robert E. Schapire (1990) ve Yoav Freund (1995) tarafından sunulmuştur. Toplulukta zayıf sınıflandırıcıların veya tahmin edicilerin bir araya getirilmesi ile güçlü bir sınıflandırıcı ya da tahmin edici oluşturmayı amaçlayan

zayıf öğrenme varsayımına dayanan aşamalı bir yaklaşımdır (Polikar, 2012).

Güçlendirme-hızlandırma algoritması hata en iyilemesine (optimizasyonu) dayalı performans artırıcı makine öğrenmesi yöntemidir. Seri içerisindeki bir model serideki bir önceki modelin tahmin hatalarını en iyilemek için kurulur (Keskin, 2019). Bu yöntem çok sayıda oluşturulan karar ağaçlarının sonuçlarını ağırlıklı oylamaya tabi tutarak son sınıf tahminini yapmaktadır. Daha önceki oluşturulan ağaçlardan yanlış tahminde bulunan tahmin edicilere daha fazla ağırlık verilerek art arda yeni ağaçlar oluşturulmaktadır.

En yaygın kullanılan güçlendirme-hızlandırma yöntemleri; Adaboost, Gradyan Arttırıcı (Gradient Boosting), Aşırı Gradyan Arttırıcı (XGBoost, Extreme Gradient Boosting), Light Gradyan Yükseltme Makinesi (LightGBM, Light Gradient Boosted Machine) ve Kategorik Gradyan Yükseltme Makinesi'dir (CatBoost, Category Gradient Boosting).

### TORBALAMA ALGORİTMASI (BAGGING)

Breiman (1996) tarafından geliştirilmiştir. Torbalama (Bagging), "bootstrap aggregation" ifadesinin kısaltılmışıdır. Önyükleme (Bootstrap) yöntemiyle mevcut veri setindeki verilerden her defasında yerine koyarak farklı örnekler seçilir. Yeni veri setleri ile çok sayıda birbirinden bağımsız karar ağaçları oluşturulur. Tahminlerin birleştirilmesi aşamasında regresyon ağaçları için ortalama alınırken sınıflandırma ağaçlarında sonuçlar oylama ile belirlenir.

Torbalama algoritması tek bir öğrenci algoritmadan elde edilen modelin performansını arttırmak için ortaya çıkan yaklaşımdır. Karar ağaçları gibi algoritmalarda yaşanan aşırı öğrenme sorununa çözüm bulmak için geliştirilen topluluk öğrenme yöntemleridir. Karar ağaçları yüksek eğitim varyansına sahiptir. Torbalanmış karar ağaçları, varyansı düşürerek nihai modeli dengeler, bu durum doğruluğu (accuracy) artırır. Sınıflandırma problemlerinde en çok oyu alan sınıfı atayarak, Regresyon problemlerinde ise her bir karar ağacının tahminlerinin ortalamasını alarak en iyi modele ulaşmaya çalışır (Keskin, 2019).

### GRADYAN ARTTIRICI MAKİNE ALGORİTMASI (GBM, GRADIENT BOOSTING MACHINE)

Jerome H. Friedman tarafından "Greedy Function Approximation: A Gradient Boosting Machine" isimli makale ile 2001 yılında tanıtılmıştır. Güçlendirme-

hızlandırma (boosting) metodolojisini kullanarak karar ağacı temelli topluluk öğrenmesi yapan makine öğrenimi algoritmasıdır. Bu yöntem, topluluktaki zayıf sınıflandırıcıların veya tahmin edicilerin (genellikle karar ağaçları) bir araya getirilmesi ile güçlü bir sınıflandırıcı ya da tahmin edici oluşturmayı amaçlayan zayıf öğrenme varsayımına dayanan aşamalı bir yaklaşımdır (Polikar, 2012). Güçlendirme algoritması hata en iyilemesine dayalı performans artırıcı makine öğrenmesi yöntemidir.

Bu algoritmada ağaçlar seri bir şekilde eğitilir. Ağaçlar birbirine bağlı olup seri içerisindeki bir model serideki bir önceki modelin tahmin veya sınıflandırma hatasını en aza indirmek için çalışır (Muratlar, 2020).

Gradyan artırıcı algoritmasının çalışma mantığı şu şekildedir; ilk olarak tahminlenecek değişkenin (hedef değişken) ortalaması alınır. Bu sayı ilk tahmin girişimi olan ilk yapraktır (Muratlar, 2020). Bu değer ile bağımlı değişken karşılaştırılarak ne kadar hatalı tahmin yapıldığı gözlemlenir. İlk ağaç, ilk yaprak sonucunda elde edilen hataları tahmin eden bir model olarak kurulacaktır. İlk tahmin ve ilk ağaçtan çıkan sonuçlar toplanırsa bağımlı değişken %100 doğru tahmin edilebilir. Ancak bu durum aşırı öğrenme olarak ifade edilir. GBM, bu sorunu aşmak için ağaçlara öğrenme oranı (learning rate) ekler. Bu değer 0 ile 1 arasında değişmektedir. Bir sonraki adım olarak doğru tahmine daha çok yaklaşabilmek için bir önceki ağacın hataları baz alınarak yeni ağaç kurulur (Singh, 2018). Ağaç eklerken kaybı en aza indirmek için meyilli azalım yöntemi kullanılır. Hatayı veya kaybı en aza indirmek için ağırlıklar güncellenir. Yeni ağaç üretimi belirlenen ağaç sayısına ulaşıncaya ya da ağaç eklenmesi anlamlı olmayıncaya kadar devam eder.

GBM üç temel unsur içerir:

1. En iyilenecek (Optimize) “Kayıp-Maliyet Fonksiyonu”
2. Tahmin yapmak için “Zayıf Öğrenici”
3. Kayıp fonksiyonunu en aza indirgeyecek zayıf öğrenicileri eklemek için bir “Katkı Modeli”

### RASTGELE ORMAN (RO) ALGORİTMASI

Breiman (2001) tarafından geliştirilen Rastgele Orman algoritması, karar ağacı gibi tek karar vericilerin bir araya getirilmesiyle oluşturulur. Bu nedenle topluluk

öğrenme yöntemleri ailesine dahil edilirler. Farklı birleştirme yöntemleri ile bir araya getirilen zayıf öğrenicilerin bir araya getirilmesi ile oluşturulan güçlü öğrenicilerin oluşturduğu topluluğun öğrenme yöntemidir. Tahmin varyansını ve yanlılığını azaltmak için tercih edilen algoritmaları içinde barındıran öğrenme yöntemidir. Birden çok modelin tahminleri kullanıldığından algoritmaların doğruluk oranı daha yüksek ve başarılıdır.

RO algoritması, karar ağacı temelli çok sayıda karar ağacı ile oluşturulan en uygun ağacı oluşturmak için kullanılan topluluk öğrenme algoritmalarından en popüler olanıdır. Breiman ve ark. (1984) tarafından geliştirilen sınıflandırma ve regresyon karar ağaçlarına (CART, Classification and Regression Trees) benzer bir şekilde ağaçlar oluşturur. Torbalama (bagging) algoritması ve öznelik torbalama olarak da adlandırılan rastgele alt uzay (RS, Random Subspace) yönteminin bileşimidir.

Ormandaki ağaçların her biri rastgele seçilmiş farklı eğitim kümelerinde ve rastgele seçilmiş özellikler ile eğitilir. Böylece aşırı uyum (overfitting) engellenir. CART algoritmalarında olduğu gibi bir düğümü iki veya daha fazla alt düğüme bölmek için “Gini indeksini”, regresyon algoritmalarında ise “Varyans Azaltma Kriterini” kullanır. Sınıflandırma problemlerinde en çok oyu alan sınıfı atayarak, regresyon problemlerinde ise her bir karar ağacının tahminlerinin ortalamasını alarak en iyi modele ulaşmaya çalışır (Keskin, 2019).

RO yönteminde, model kurulurken, modeli test etmek için ayrı bir test veri seti yoksa veya orijinal veri setinden test veri seti ayrılmamışsa, sınıf dağılımına bağlı kalınarak orijinal veri setinin  $2/3$ ’ü öğrenme veri seti (inBag),  $1/3$ ’ü ise test veri seti (Out-Of-Bag (OOB)) olarak ayrılmaktadır. Eğer ayrı bir test veri seti varsa veya orijinal veri setinden test veri seti ayrılmışsa, modelin kurulması için ayrılan öğrenme veri seti, kendi içinde  $2/3$  oranında öğrenme veri seti (inBag),  $1/3$ ’ü ise test veri seti (OOB) olarak ayrılmaktadır. Karar ormanı kaç karar ağacı ile oluşturulacaksa o kadar sayıda önyükleme (bootstrap) tekniği ile örneklem oluşturulur ve her örneklem için inBag ve OOB verisi ayrılır. Kurulan ağacın, OOB verisi ile testi yapılarak hata oranı tahmini yapılır. Tüm ağaçlar için OOB hata oranlarının ortalaması alınarak, modelin OOB hata oranı kestirimi hesaplanır. OOB verisi ile yapılan teste modelin iç testi de denilmektedir. Modelin testi, ayrı bir test veri seti ile de yapılabilir (Akman, 2010).

Rastgele seçilen değişkenlerden en iyi bölünme



sağlayacak olan belirlenip, dallara ayrılmaya o değişkenden başlanılır. Uygulamada analistin belirlemesi gereken iki parametre vardır, bunlar oluşturulacak ağaç sayısı ve seçilecek değişken sayısıdır. Bununla birlikte, bu parametrelerin seçilecek değerlerin genellikle sonuca etkisi azdır (Akman, 2010).

## ALGORİTMA UYGULANMA ve PERFORMANS DEĞERLENDİRİLME AŞAMALARI

### VERİ ÖN İŞLEME SÜRECİ

Makine öğrenmesi algoritmaları büyük verilerin işlenmesi için geliştirilmiştir. Büyük veri, daha fazla çeşitlilik içeren ve hacmi hızlıca artan özellikle yeni veri kaynaklarından elde edilen daha büyük, daha karmaşık veri kümeleridir. Bu veri kümeleri içinde metin, ses ve video gibi yapılandırılmamış ve yarı yapılandırılmış veri türleri bulunduğu için bunlardan anlam türetmek ve meta verileri desteklemek için veri ön işleme gereklidir.

Verilerin önceden işlenmesi, makine öğrenme algoritmaları üzerindeki verileri değerlendirmeden önce iyi bir tahmin performansı elde etmek için önemli bir adımdır. Verilerden aykırı değerlerin silinmesi, eksik verilerle ilgilenilmesi, veri normalizasyonu ve ağırlıklandırma ile özellik (öznitelik) seçme işlemi gibi birtakım süreçlerden geçirilerek veri ön işleme gerçekleştirilmiş olur (Akay, 2018).

Veri normalizasyonu, birbirlerinden farklı birimlerle ifade edilen sayısal değerlerin ortak bir aralıkta dönüşümünü içeren önemli bir ön işleme aşamasıdır. Normalleştirme ham verilerin üzerinde, her bir özelliğin eşit bir katkısı olacak şekilde yeniden ölçeklendiren veya dönüştüren bir işlemdir. Tüm özellikler öğrenme sürecine eşit katkıda bulunur. Ham (normalleştirilmemiş) verilerden elde edilen istatistiksel önlemlere dayanarak verileri belirli bir aralıkta normalleştirmek için iki adet yöntemler ailesi önerilmiştir (Öztürk, 2020).

İlki ortalama ve standart sapmaya dayalı normalleştirme yöntemidir. Verileri normalleştirmek için ham verilerin istatistiksel ortalaması ve standart sapması kullanılır. Bu yaklaşımı kullanan birçok farklı yöntem bulunmaktadır. Bunlar; merkezi ortalama,

Pareto ölçeklendirme, değişken kararlılık ölçekleme, güç dönüşümü ve z-değeri normalizasyonudur.

İkincisi, en küçük-en büyük (min-max) değer tabanlı normalleştirme yöntemidir. Normalleştirilmemiş verilerin en küçük ve / veya en büyük değerleri yeniden ölçeklendirme için kullanılır. Bu yöntemler şunlardır; en küçük-en büyük normalleştirme ve en büyük normalleştirme.

Veri ağırlıklandırma, özellik uzayındaki dağılımları doğrusal olarak ayırlamayan veri kümelerinin doğrusal olarak ayrılabilir hale getirilmesidir. Veri kümesi içerisindeki özelliklerin varyansını azaltarak bunu sağlar. Veri ağırlıklandırma yöntemlerine örnek olarak ortalama kayma kümeleme tabanlı ve ikili bulanık c-ortalama tabanlı veri ağırlıklandırma verilebilir (Singh, 2018).

### ÖZİNTELİK SEÇİM YÖNTEMLERİ

Özellik seçme işleminde veri kümelerindeki özelliklerden önem derecesi yüksek olanlar seçilir. Bir boyut azaltma işlemidir. Gereğinden fazla değişkenin olduğu bir modelin açıklanması ve yorumlanması oldukça zordur. Bu yüzden öznitelik seçimi ile modeli daha kolay anlaşılabilir hale dönüştürebiliriz.

Öznitelik seçimi ile modelde oluşabilecek aşırı öğrenme durumunun önüne geçilebilir. Modelin doğruluk oranı artmış olur. Bir veri setinde değişkenler arasında yüksek ilişki yani çoklu bağlantı (multicollinearity) olabilir. Eğitim aşaması hızlanmış olur. Model başarı oranları artırılabilir (Ünal, 2015).

### HİPER PARAMETRE AYARLAMASI (TUNİNG)

Hiper parametreler modelin sihirli sayılarıdır. Algoritmanın performansını ayarlayabilen parametrelerdir. Model eğitime başlamadan önce kullanıcı tarafından karar verilen değerlerdir. Hiper parametre ayarlaması modellemenin en önemli aşamalarından biridir. Hiper parametre değerlerinde yapılan en ufak değişiklikler model performansında çok büyük etkiler oluşturur. Yapay sinir ağı algoritmaları için öğrenme oranı, momentum oranı, gizli katman sayısı, eğitim tur (epoch) sayısı ve gizli katmandaki nöron sayısı hiper parametrelere örnek olarak verilebilir (Johnson, 2017).

Model tasarımı yapılırken hiper parametrelere yönelik yaptığımız ilk seçimler genelde bizi doğru sonuçlara götürmez. İteratif bir şekilde ardı ardına hiper parametre değerleri değiştirilerek, modelin başarımı gözlenir ve model için en uygun hiper parametre grubu seçilmeye çalışılır. Bu

parametreler için en uygun değerleri bulma yöntemleri; sezgisel yolla, ızgara arama (grid search) ve rastgele arama ile parametre uydurmadır (Çarkacı, 2018).

Sezgisel parametre uydurma yöntemi ile probleme dair ön bilgilerimizi kullanarak hiper parametreler tahmin edilir, model bu hiper parametrelere göre tasarlanır ve sonuçlar gözlenir. Çıkan sonuçlara göre modelin başarısını artıracak yeni değerler ile model tekrar oluşturulup, eğitilir ve sonuçlar gözlenir (Çarkacı, 2018).

Izgara arama ile uygun parametre bulma yönteminde hiper parametrelerin bazıları sonsuz sayıda değer alabilecek konumda olduğu için biz problem hakkında sahip olduğumuz ön bilgileri kullanarak değer aralıkları belirleriz. Bu aralıklardan belirli ana noktalar seçilerek hiper parametreler için değer listeleri oluşturulur. Belirlenen aralıkta bulunan tüm değerlerin kombinasyonları için ağ eğitilip sonuçlar gözlenir duruma göre en iyi kombinasyon hiper parametre grubu olarak seçilir (Çarkacı, 2018).

Rastgele arama ile uygun parametreleri bulma yönteminde ızgara aramada olduğu gibi probleme dair ön bilgiler kullanılarak hiper parametre aralıkları belirlenir. Daha sonra bu aralıktaki değerlerin her birini denemek yerine rastgele değerler seçilerek hiper parametre grupları oluşturulur. En iyi sonucu bulana kadar devam edilir (Çarkacı, 2018).

## MODEL PERFORMANS DEĞERLENDİRME YÖNTEMLERİ

Model performans değerlendirme yöntemleri mevcut veri setinin eğitim ve test veri seti olarak nasıl bölüneceğini örneklemelerin nasıl elde edileceğini gösterir (Arat, 2021). Problemin türüne göre seçilen makine öğrenmesi algoritmaları ile veri setinden eğitim için ayrılan veriler ile öğrenilen model deseni elde edilir ve test veri setiyle de modelin performansı değerlendirilir.

Model performans değerlendirme yöntemleri şunlardır; dışarıda tutma (holdout), tekrarlı dışarıda tutma (repeated holdout), tabakalı örnekleme, üçlü ayırma, k-katmanlı çapraz doğrulama (k-fold cross validation), rastgele örnekleme yöntemi ve önyüklemeli (bootstrap) örneklemedir (Arat, 2021).

Makine öğrenme algoritmalarının eğitim veri

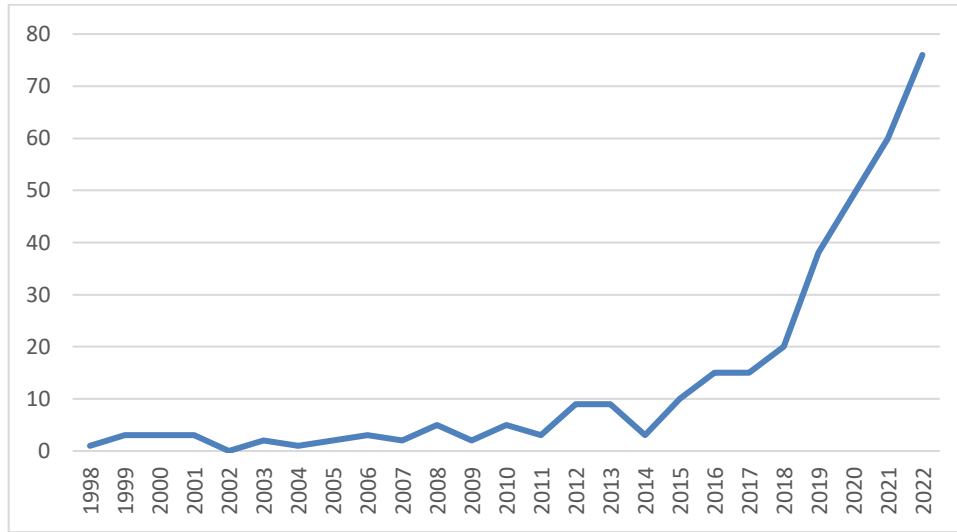
setinden öğrendiği modeli test etmek için kullanılan en yaygın model performans değerlendirme yöntemi k-katmanlı çapraz doğrulamadır. Çapraz doğrulama, yapılan bir istatistiksel analizin bağımsız bir veri setinde nasıl bir sonuç elde edeceğini sınavan bir model doğrulama tekniğidir (Karlı, 2019). Modelin doğru sınıflama performansını bağımsız bir veri setine gerek olmadan test etmek için kullanılan bir yöntemdir. Veri seti rastgele olarak k adet parçaya bölünür. Nedeni birbirine benzer homojen dağılım gösteren alt kümeler elde etmektir. K adet parçadan k-1 tanesi eğitim veri seti olarak birleştirilir. Geriye kalan parça ise test veri seti olarak kullanılır. K defa tekrarlanan bu süreçte her bir tekrarda k-1 adet eğitim veri setinden algoritma bir model öğrenir ve bu modelin tahmin ve doğruluk oranları test veri seti üzerinden sınanır (Özgür, 2020).

## MODEL PERFORMANS DEĞERLENDİRME ÖLÇÜTLERİ

Makine öğrenmesi algoritmaları kullanılarak regresyon problemi çözümünün başarılı olup olmadığını değerlendirmek için performans değerlendirme kriterleri olarak bazı istatistiksel hata ölçütleri kullanılmaktadır. Bu ölçütler, ortalama mutlak yüzde hatası (MAPE, Mean Absolute Percentage Error), ortalama mutlak hata (MAE, Mean Absolute Error), ortalama kare tahmin hatası (MSE, Mean Square Error) ve ortalama tahmin hatasının karesinin karekökü (RMSE, Root Mean Square Error) olarak sıralanabilirler (Kwon, 2011).

## VETERİNER HEKİMLİĞİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK YAPILAN ÇALIŞMALAR

Veteriner hekimliği alanında makine öğrenmesi ile ilgili yayınlar her geçen gün artmaktadır. WOS'ta "Veteriner hekimlik" ve "Makine öğrenmesi" anahtar kelimeleri ile literatür taranmış 1998 ve 2022 yılları arasında toplam 362 makalenin yayınlanmış olduğu görülmüştür (Grafik 1.). Bunların 316'sı araştırma makalesi ve 46'sı derlemedir. Bu çalışmada, konuyla ilgili kapsamlı bir inceleme makalesi sağlamaktan ziyade makine öğrenmesi algoritmalarının veteriner hekimliği alanındaki potansiyelini göstermek amacıyla çeşitli örnekler derlenmiştir. Bu noktada, beşeri hekimlikte ML yaklaşımları ve veri bilimi hakkında daha fazla ayrıntı sağlayan pek çok araştırma makalesi de bulunmaktadır.



Grafik 1. 1998-2022 tarihleri arası yayınlanan makaleler

Aşağıdaki Tablo 1.'de veteriner hekimliği alanında makine öğrenmesi regresyon algoritmaları ile yapılan çalışma örnekleri gösterilmiştir.

Tablo 1. Regresyon algoritmalarının kullanıldığı bazı araştırma makaleleri

Yazar	Yıl	n	Metot (Algoritmalar)	Amaç
Takma ve arkadaşları	2012	305	Çoklu doğrusal regresyon ve yapay sinir ağı (YSA)	Siyah Alaca ineklerin laktasyon süt verimi tahmini
<b>Sonuç;</b> analizler sonucunda yapay sinir ağları modeli ile ilk beş laktasyona ait belirleme katsayı ( $R^2$ ) değerleri 0,62-0,85 arasında, hata kareler ortalamasının karekökü (RMSE) değerleri 480,9-1682,8 arasında, ortalama mutlak sapma (MAD) değerleri 325,2-1381,7 ve ortalama mutlak yüzde hata (MAPE) değerleri 6,1-20,2 arasında değişim göstermiştir. Çoklu doğrusal regresyonda ise bu değerler sırası ile $R^2$ için 0,30 ile 0,75 arasında, RMSE için 1964,8-3008,7 arasında, MAD değeri 1576,6-2458,3 arasında ve MAPE değeri 24,7-35,6 arasında bulunmuştur. Söz konusu kriterlere göre yapay sinir ağı modelinin çoklu doğrusal regresyon modelinden daha iyi uyum sağladığı gözlenmiştir.				
Ghazanfari	2014	72	Çok Katmanlı Algılayıcı (Multilayer Perceptron-MLP) tipi yapay sinir ağları geriye yayılım algoritması, Doğrusal Regresyon	Ağırlık artışına, yem tüketimi (feed intake) ve yem dönüşüm oranına (feed conversion ratio) etkisi sonuç değişkenleri olarak kullanılmıştır.
<b>Sonuç;</b> yapay sinir ağından elde edilen modelin eğitim ve test değerleri doğrusal regresyona göre daha iyi bulunmuştur.				
Akkol ve arkadaşlarının	2017	475	Yapay Sinir Ağları ve Çoklu doğrusal regresyon analizi	Kıl keçisine ait Morfolojik özellik ölçümlerinin canlı ağırlık üzerine etkileri
<b>Sonuç;</b> Levenberg-Marquart, Bayes düzeltmeli (Bayesian regularization) ve Scaled conjugate olmak üzere üç farklı geri yayılım algoritması kullanılmış Bayes düzeltmeli geriye yayılım algoritması kullanan yapay sinir ağlarının çoklu doğrusal regresyon analizine göre daha başarılı olduğu belirlenmiştir.				
Takma ve Gevrekçi	2018		Yapay Sinir Ağı	Yumurta verimi tahmini
<b>Sonuç;</b> ileri geribildirim algoritması tarafından eğitilen çok tabakalı yapay sinir ağı modeli, ileri beslemeli sinir ağı öğrenmesi için kullanılmıştır. Eğitim ve test seti için belirleme katsayı ( $R^2$ ) değeri sırasıyla 0,80 ve 0,82 olarak tahminlenmiştir. Çalışmada düşük hata kareler ortalamasının karekökü (RMSE) ve ortalama mutlak sapma (MAD) değerleri elde edilmiştir. Uygulama sonucuna göre yumurta üretiminin tahmininde yapay sinir ağı kullanılabilirliği belirlenmiştir.				

Tablo 1. Devamı.

Yazar	Yıl	n	Metot (Algoritmalar)	Amaç
Nguyen ve arkadaşları	2020	36	Çoklu doğrusal regresyon, Rastgele Orman, Destek Vektör Makinesi ve Yapay Sinir Ağı (YSA)	İnek süt verimi tahmini
<b>Sonuç;</b> analiz sonucunda Destek Vektör Makinesi ile elde edilen regresyon modelinin, süt verimine ilişkin en doğru tahmini en kısa sürede verdiği belirlenmiştir.				
Chen ve ark.	2022	951	Çoklu doğrusal regresyon (ÇDR), Yapay Sinir Ağı (YSA) Rastgele Orman (RO) ve Destek Vektör Makinesi (DVM)	Emziren süt ineklerinde gübre azotu atılımını tahmin etmek için çoklu doğrusal regresyon ve makine öğrenme algoritmalarının performansını karşılaştırmayı ve Gübre azotu atılımı için yeni makine öğrenimi tahmin modelleri geliştirmeyi amaçlamaktadır.
<b>Sonuç;</b> Gübredeki azotun tahmini için önerilen yeni YSA modeli on kat çapraz doğrulamada ÇDR, RO ve DVM modeline kıyasla daha düşük bir RMSE ve daha yüksek bir Konkordans uyum katsayısı (CCC) üretmiştir. YSA modellerinin performansının, özelliklerin seçimi ve öğrenme algoritmalarının döndürülmesi süreciyle büyük ölçüde iyileştirildiğini göstermiştir.				
Weleszczuk ve ark.	2022	2000	Çoklu doğrusal regresyon (ÇDR), Karar Ağacı, Yapay Sinir Ağı (YSA) Rastgele Orman (RO), K-en yakın komşunun (KNN), Gradyan Arttırıcı Makine (GAM), Aşırı gradyan arttırıcı regresyon (XGBoost)	Polonya Holstein sığırları için iki önemli ekonomik faktörü öngöreceği yeni bir yöntem önermektedir. Birincisi ekonomik endekstir (EI), İkinci faktör, doğum yapmak ile tarafından başka bir gebeliğe girmek arasındaki zaman dilimi olarak kabul edilen buzağılama aralığıdır (CI).
<b>Sonuç;</b> Elde edilen ortalama mutlak sapma (MAD) ve hata kareler ortalamasının karekökü (RMSE) sonuçları, EI'yi tahmin etmek için en iyi modelin YSA'ya dayalı olarak oluşturulan model ve GAM tabanlı CI modeli için oluşturulan model olduğunu göstermiştir. Belirleme katsayısı ve RMSE sonuçlarına dayanarak, önerilen XGBoost modelinin diğer modellerden daha iyi performans gösterdiği ortaya çıktı.				

RMSE: Root Mean Square Error, MAD: Mean Absolute Deviation, MAPE: Mean Absolute Percentage Error, CCC: Concordance correlation coefficient.

Tablo1.'deki çalışmalara ek olarak Slob ve ark. (2020) tarafından yapılan sistematik derlemede, mandıra sektöründe makine öğrenmesi uygulamaları Scopus, Web of Science, IEEE, Wiley, Springer Link veri tabanlarında taranmış, 427 makaleye ulaşılmış 98 tekrar eden çalışma çıkarılarak 329 çalışma elde edilmiştir. Dahil etme ve çıkartma kriterlerine göre 2010 ile 2020 tarih aralığında 38 çalışma araştırmaya dahil edilmiştir. Bunların %55'i hastalık araştırması, %26'sı ürün ve %19'u kalite tahmin araştırmasıdır. Meme iltihabı (mastitis) tahmini en çok kullanılan bağımlı değişken olmuş (%50). Diğer çok kullanılan değişkenler ise süt verimi ve kalitesidir. En çok kullanılan yöntemler sırasıyla, Ağaç temelli algoritmalar ve Yapay sinir ağlarıdır. 23 farklı algoritma kullanılmıştır. Performans değerlendirme ölçütü olarak %50'den fazla çalışmada duyarlılık, seçicilik ve RMSE kullanılmıştır. En çok sınıflandırma problemleri için sonra sırasıyla regresyon ve kümeleme problemleri için algoritmalar kullanılmıştır. 21 çalışmada dışarda tutma yöntemi ile 14 çalışmada ise

k-katmanlı çapraz doğrulama yöntemi ile eğitim ve test veri seti oluşturulmuştur. 17 çalışmada en iyi performansı ANN ve 15 çalışmada ise ağaç tabanlı algoritmalar göstermiştir.

## SONUÇ

Yapay zekanın diğer alanlarda olduğu gibi veteriner hekimlikte uygulanma amacı, farklı kaynaklardan elde edilen farklı yapıdaki (yapılandırılmış, yapılandırılmamış ve yarı yapılandırılmış) büyük ve daha karmaşık veri kümelerinden geleneksel istatistiksel yöntemler ile elde edilemeyen desenlerin ortaya çıkarılabilmesidir.

Veterinerlik alanında da artan veri sayısı ve büyük verilerin karmaşıklığı, makine öğrenimi ihtiyacını artırmıştır. Chen ve He (2014) makine öğrenmesini genellikle geleneksel veri işleme araçları ile analizi yapılamayan ve karmaşık bir yapıya sahip büyük miktardaki veri setlerini kullanarak anlamlı ilişkiler bulan model ve desenler geliştiren yöntemler ailesi olarak tanımlamışlardır. Veterinerlik alanındaki

sorunlar için farklı çözüm yolları teknolojinin yeni geliştirdiği araçlar ile bulunabilir.

Bu derleme çalışmasında literatürde çok sık kullanılan makine öğrenmesi regresyon algoritmalarından, veri ön işleme aşamalarından, özellik ya da öznelik seçme yöntemlerinden, hiper parametre en iyileme yöntemlerinden, model performans değerlendirme yöntemlerinden ve model performans değerlendirme ölçütlerinden kısaca bahsedilmiştir.

Çalışmada ilk olarak denetimli öğrenme yöntemi ile eğitilen makine öğrenmesi regresyon algoritmaları hakkında kısa bilgiler verilmiştir. Literatürde tahmin problemlerinin çözümünde sıklıkla kullanılan klasik makine öğrenmesi regresyon yöntemleri olan doğrusal regresyon, Ridge, LASSO ve Elastik Net düzenleme algoritmaları ile hem sınıflandırma hem de regresyon problemleri için kullanılabilecek bulunan YSA, Naïve Bayes, karar ağaçları, kNN, DVM ve DVR algoritmaları hakkında tarihsel gelişim sürecine göre genel bilgiler sunulmuştur. Sonraki aşamada ise günümüzde yaygın olarak kullanılan topluluk öğrenme yöntemleri olan Torbalama, Güçlendirme-Hızlandırma, Gradyan Arttırıcı ve Rastgele Orman algoritmaları tanıtılmıştır.

Çalışmanın ikinci bölümünde veri ön işleme sürecinde yapılan işlemler ve kullanılan yöntemler kısaca açıklanmıştır. Makine öğrenmesi regresyon algoritmalarının uygulanması sırasında amaca uygun bir biçimde toplanan verilerden aykırı değerlerin silinmesi, eksik verilerle ilgilenilmesi, veri normalizasyonu ve ağırlıklandırma gibi süreçlerden kısaca bahsedilmiştir. Bu aşamada yapılan her bir işlem için sunulan farklı yöntem (algoritma) seçeneklerinden en uygun olanı kullanılarak oluşturulan algoritma kombinasyonlarının uygunluğu ve doğruluğunun makine öğrenmesi regresyon algoritmalarının daha hızlı eğitilmesine olanak sağlayacağı belirtilmiştir.

İkinci bölümün devamında, seçim yöntemlerinden bahsedilmiş gereğinden fazla özneliğin olduğu bir modelin açıklanması ve yorumlanmasının oldukça zor olduğu; modelde oluşabilecek aşırı öğrenme durumunun önüne geçebilmek, eğitim aşamasını hızlandırmak ve model başarı oranlarını arttırmak için öznelik seçiminin

önemi anlatılmıştır.

Sonrasında; algoritmanın performansını ayarlayabilen, modelin performansını doğrudan etkileyebilen ve modelin sihirli sayıları olarak bilinen hiper parametrelerin ayarlanmasından kısaca bahsedilmiştir. Bu aşamanın modellemenin en önemli bölümlerinden biri olduğu, en ufak değişikliğin model performansında yarattığı etki ve en uygun hiper parametre değerleri bulununcaya kadar iteratif olarak algoritma eğitiminin devam edeceği vurgulanmıştır. Çarkacı (2018) "Sezgisel yolla", "Izgara arama" ve "Rastgele arama ile parametre uydurma" gibi isimler ile üç farklı yöntem (algoritma) grubundan en uygun olanın seçiminin modelin performansını artıracakını ifade etmiştir.

Bir sonraki aşamada veri setinin eğitim ve test veri seti olarak nasıl ayrılacağına karar vermek için kullanılan model performans değerlendirme yöntemlerinden bahsedilmiştir. Dışarıda tutma, tekrarlı dışarıda tutma, üçlü ayırma, k-katmanlı çapraz doğrulama, tabakalı, rastgele ve önyüklemeli örnekleme gibi yöntemlerin kullanıldığı ve bunlardan en yaygın kullanılan yöntemin k-katmanlı çapraz doğrulama olduğu belirtilmiştir. Eğitim ve test veri setlerini ayırmada kullanılan performans değerlendirme ölçütlerinin analiz sonuçlarına doğrudan etki edeceği belirtilmiştir.

İkinci kısmın son bölümünde algoritmanın eğitimi ile elde edilen modelin ve bu modelin test veri seti ile sınanması ile elde edilen model performans değerlendirme ölçütlerinden bahsedilmiştir. Regresyon problemi çözümünün başarısını değerlendirmek için kullanılan istatistiksel hata ölçütleri olan MAPE, MAE, MSE ve RMSE'nin en sık kullanılan değerlendirme kriterleri olduğu Kwon (2011) tarafından belirtilmiştir.

Üçüncü kısımda makine öğrenmesi regresyon algoritmalarının veterinerlik alanında uygulandığı çalışmalar incelenmiştir.

Sonuç olarak, hızla gelişen ve kendini çok hızlı yenileyen bu teknolojik sürece veteriner hekimlerin uzak kalması beklenemez. Veteriner hekimlerin ilerleyen yıllarda makine öğrenmesi algoritmaları ile elde edilen sonuçların sahaya uygulanması aşamasında en son karar verici olarak yapılan işlemleri çok iyi bilmeleri gerekliliği

görülmektedir.

Hayvan hastalıklarının teşhis, tanı ve tedavisinde başarılı küçük yapay zekalar geliştirilebilmek ve veteriner hekimlik alanında kullanılacak robotik zekaların bileşenlerinden olacak bu küçük yapay zekaların geliştirilmesinin multidisipliner bir alan olduğu görülmektedir. Bu yüzden, örneğin, veterinerlik alanında görev yapan patologların yapay zeka çalışmalarından en azami şekilde faydalanabilmeleri için klinisyen, moleküler biyolog ve biyomedikal mühendislerin yanında imaj ve bilgisayar bilimciler ile biyoinformatik uzmanlarına ihtiyaç duyulacağı La Perle (2019) tarafından öngörülmektedir.

Hayvan sağlığı için kullanılacak ve veteriner hekimlerin işini kolaylaştırmak için geliştirilecek Tele tıp teknolojisinin gelişim aşamasında veteriner hekimlerin yapay zekanın alt dalı olan makine öğrenmesi regresyon algoritmaları ve uygulamaları sırasında dikkat etmeleri gereken hususlar hakkında bilgi sahibi olmaları gereği gün geçtikçe artmaktadır.

Makine öğrenmesi tekniklerinin veteriner hekimlik araştırmalarında kullanılması, ileriye dönük tahminlerde bulunma ve istenmeyen durumlara karşı tedbir alma açılarından veteriner hekimlere büyük fırsatlar sunmaktadır. Veteriner hekimin işini kolaylaştırmak için veterinerlik alanında Makine öğrenmesi regresyon yöntemlerinin kullanımı oldukça önem arz etmektedir.

#### KAYNAKÇA

- Akay, E. Ç. (2018). Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. *Sosyal Bilimler Araştırma Dergisi*, 7(2), 41-53.
- Akkol, S., Akıllı, A. & Cemal, İ. (2017). Kıl Keçilerinin Canlı Ağırlık Tahmininde Yapay Sinir Ağları ve Çoklu Doğrusal Regresyon Yöntemlerinin Karşılaştırılması. *Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi*, 27, 21-29.
- Akman, M. (2010). Veri Madenciliğine Genel Bakış ve Random Forest Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama. AÜ Sağlık Bilimleri Enstitüsü Biyoistatistik Ana Bilim Dalı Yüksek Lisans Tezi, Ankara.
- Alpar, R. (2011). Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. 3üncü Baskı. Ankara: Detay Yayıncılık.
- Arat, B. (2021). Makine Öğrenmesi Model Performans Değerlendirme Yöntemleri. Erişim Adresi: <https://berkarat.com/model-performans-degerlendirme-yontemleri/> Erişim Tarihi: 24.04.2021.
- Atalay, M. & Celik, E. (2017). Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları - Artificial Intelligence and Machine Learning Applications in Big Data Analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9, 155-172.
- Balta, A. (2018). Makine Öğrenmesi Teknikleri İle Ekolojik Verilerin Değerlendirilmesi. FÜ Fen Bilimleri Enstitüsü Ekobilişim Ana Bilim Dalı Yüksek Lisans Tezi, Elazığ.
- Basran, P. S. & Appleby, R. B. (2022). The unmet potential of artificial intelligence in veterinary medicine. *American Journal of Veterinary Research*, 83, 385-392.
- Bhattacharyya, S. (2018). Ridge and Lasso Regression: L1 and L2 Regularization. Erişim Adresi: <https://medium.com/towards-data-science/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b/> Erişim Tarihi: 10.05.2023.
- Boser, E., Guyon, I. M. & Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. In D. Haussler (Ed.), Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (pp. 144-152). Pittsburgh, PA: ACM Press.
- Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). Classification and Regression Tree. 1st ed. New York: Routledge.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123-140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Bulut, M. A. (2019). Kredi Analizinde Makine Öğrenmesi Kullanımı: Tarımsal Kredilerde Uygulama Örneği. OÜ Sosyal Bilimler Enstitüsü İşletme Ana Bilim Dalı İşletme Bilim Dalı Doktora Tezi, Eskişehir.
- Ceylan, T. (2018). Perakende Sektöründe Makine Öğrenmesine Dayalı Yaklaşımlar. YTÜ Fen Bilimleri Enstitüsü İstatistik Ana Bilim Dalı İstatistik Bilim Dalı Yüksek Lisans Tezi, İstanbul.

- Chen, T. & He, T. (2014). Higgs Boson Discovery with Boosted Trees. Proceedings of the International Conference on High-Energy Physics and Machine Learning, Montreal, Canada, 2014, 69–80.
- Chen, X., Zheng, H., Wang, H. & Yan, T. (2022). Can machine learning algorithms perform better than multiple linear regression in predicting nitrogen excretion from lactating dairy cows. *Sci Rep*, 12, 12478.
- Cihan, P., Gökçe, E. & Kalıpsız, O. (2017). A Review of Machine Learning Applications in Veterinary Field. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 23, 673-680.
- Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13, 21-27.
- Çağlayan Akay, E. (2018). Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. *Sosyal Bilimler Araştırma Dergisi*, 7, 41–53.
- Çarkacı, N. (2018). Derin Öğrenme Uygulamalarında Hiper Parametre Seçim Yöntemleri. Erişim Adresi: <https://medium.com/deep-learning-turkiye/derin-ogrenme-uygulamalarinda-model-dogrulama-ve-hiper-parametre-secim-yontemleri-823812d95f3/> Erişim Tarihi: 15.05.2021.
- Freund, Y. (1990). Boosting a weak learning algorithm by majority. In: Proceedings of the third annual workshop on computational learning theory, Morgan-Kaufmann, 1990, 202–216.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.
- Gandhi, R. (2018). Introduction to Machine Learning Algorithms: Linear Regression. Erişim Adresi: <https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a> Erişim Tarihi: 21.12.2020.
- Ghazanfari, S. (2014). Application of linear regression and artificial neural network for broiler chicken growth performance prediction. *Iranian Journal of Applied Animal Science*, 4, 411-416.
- Hoerl, A. E. & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 55–67.
- Jason, B. (2019). A Tour of Machine Learning Algorithms. Erişim Adresi: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/> Erişim Tarihi: 20.03.2022.
- Johnson, A. (2017). Common Problems in Hyperparameter Optimization. Erişim Adresi: <https://sigopt.com/blog/common-problems-in-hyperparameter-optimization> Erişim Tarihi: 03.04.2021.
- James, G., Eitten, D., Hastie, T. & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R, 4th ed. New York, USA: Springer Science+Business Media.
- Karlı, Ö. B. (2019). Makine Öğrenme Yöntemleri ile Karaciğer Hastalığının Teşhisi. İÇÜ Fen Bilimleri Enstitüsü Matematik Ana Bilim Dalı Bilgisayar Bilimleri Bilim Dalı Yüksek Lisans Tezi, Ağrı.
- Keskin, M. V. (2019). Ağaca Dayalı Yöntemlerde Bagging ve Boosting Arasında Ne Fark Var? Erişim Adresi: <https://www.veribilimiokulu.com/agaca-dayali-yontemlerde-bagging-ve-boosting-arasinda-ne-fark-var/> Erişim Tarihi: 13.12.2020.
- Kıral Özkan, T. (2015). Makine Öğrenmesi Yöntemleri ile Vakıf Üniversiteleri Doluluk Tahminlemesi. MÜ Sosyal Bilimler Enstitüsü Ekonometri Ana Bilim Dalı Yöneylem Araştırması Bilim Dalı Doktora Tezi, İstanbul.
- Kwon, S. J. (2011). Artificial Neural Networks, UK ed. UK: Nova Science Publishers, Inc.
- La Perle, KMD. (2019). Machine Learning and Veterinary Pathology: Be Not Afraid!. *Veterinary Pathology*, 56, 506–507.
- Maron, M. E. (1961). Automatic indexing: an experimental inquiry. *JACM*, 8, 404-417.
- McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5, 115-133.
- Muratlar, E. R. (2020). Gradient Boosted Regresyon Ağaçları. Erişim Adresi: <https://www.veribilimiokulu.com/gradient-boosted-regresyon-agaclari/> Erişim Tarihi: 09.04.2021.
- Morgan, J. N. & Sonquist, J. A. (1963). Some results from a non-symmetrical branching process that looks for interaction effects. *Young*, 8.
- Nguyen, Q. T., Fouchereau, R., Frénod, E., Gerard, C. & Sincholle, V. (2020). Comparison of forecast models

- of production of dairy cows combining animal and diet parameters. *Comput Electron Agric*, 170, 105258.
- Özgür, E. G. (2020). Rasch Modeline Uygun Yanıt Desenlerinde Bilgisayar Uyarlamalı Test Yöntemi ile Makine Öğrenmesi Yöntemlerinin Tahmin Performanslarının Benzetim Çalışması İle Karşılaştırılması. AÜ Sağlık Bilimleri Enstitüsü Biyoistatistik Ana Bilim Dalı Doktora Tezi, Ankara.
- Öztürk, S. (2020). Hızlı Moda Sektöründe Makine Öğrenmesi Yöntemleri ile Satış Miktarlarının Tahmin Edilmesi. İTÜ Fen Bilimleri Enstitüsü Endüstri Mühendisliği Ana Bilim Dalı Endüstri Mühendisliği Bilim Dalı Yüksek Lisans Tezi, İstanbul.
- Polikar, R. (2012). Ensemble Learning. In, Zhang C., Ma Y. Editors. Ensemble Machine Learning: Methods and Applications. Boston, MA, USA: Springer; 2012. pp. 1-34.
- Salunke, D. (2023). Implementing SVM and Kernel SVM with Python's Scikit-Learn. Erişim Adresi: <https://www.geeksforgeeks.org/implementing-svm-and-kernel-svm-with-pythons-scikit-learn/> Erişim Tarihi: 21.11.2023.
- Schapiro, R. E. (1990). The Strength of Weak Learnability. *Machine Learning*, 5, 197-227.
- Singh, H. (2018). Understanding Gradient Boosting Machines. Erişim Adresi: <https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab/> Erişim Tarihi: 12.07.2021.
- Slob, N., Catal, C. & Kassahun, A. (2020). Application of Machine Learning to Improve Dairy Farm Management: A Systematic Literature Review. *Preventive Veterinary Medicine*, 187, 105237.
- Smola, A. & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199-222.
- Şirin, E. (2017). Karar Ağacı ile Regresyon (Decision Tree Regression): Python Örnek Uygulama. Erişim Adresi: <https://www.veribilimiokulu.com/karar-agaci-ile-regresyon-decision-tree-regression-python-ornek-uygulama/#:~:text=Karar%20ağaçlarını%20sınıflandırmaya%20ve%20regresyon,eğitim%20esnasında%20öğrendiği%20ortalamay%20söyleyiyor/>
- Erişim Tarihi: 22.06.2022.
- Ünal, Y. (2015). Makine Öğrenmesi Yöntemleriyle Bel Bölgesi Rahatsızlıklarının Tanısı. SÜ Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalı Doktora Tezi, Konya.
- Yılmaz, H. (2014). Random Forests Yönteminde Kayıp Veri Probleminin İncelenmesi ve Sağlık Alanında Bir Uygulama. OÜ Sağlık Bilimleri Enstitüsü Biyoistatistik Ana Bilim Dalı Yüksek Lisans Tezi, Eskişehir.
- Takma, Ç., Atıl, H. & Aksakal, V. (2012). Çoklu doğrusal regresyon ve yapay sinir ağı modellerinin laktasyon süt verimlerine uyum yeteneklerinin karşılaştırılması. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 18, 941-944.
- Takma, Ç. & Gevrekçi, Y. (2018). Use of neural network model to predict of egg yield. *Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi*, 35, 147-151.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267-288.
- Trafalis, T.B. & Ince, H. (2000). Support vector machine for regression and applications to financial forecasting. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, 6, 348-353 vol.6.
- Wang, Y., Wang, B. & Zhang, X. (2012). A new application of the support vector regression on the construction of financial conditions index to CPI prediction. *Proc Comput Sci*, 9, 1263-1272.
- Weleszczuk, J., Kosinska-Selbi, B. & Cholewińska, P. (2022). Prediction of Polish Holstein's economical index and calving interval using machine learning. *Livestock Science*, 265, 105039.
- Zou, H. & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 67, 301-320.