



Araştırma Makalesi (Research Article)

**ŞARAP KALİTESİNİN BELİRLENMESİNDE YAPAY ZEKÂNIN ROLÜ: MAKİNE
ÖĞRENMESİ ALGORİTMALARI ÜZERİNE BİR UYGULAMA (THE ROLE OF ARTIFICIAL
INTELLIGENCE IN DETERMINING THE WINE QUALITY: AN APPLICATION ON MACHINE
LEARNING ALGORITHMS)**

Beybala TİMUR^{1*} (orcid.org/ 0000-0002-7514-1129)

Veysel YILMAZ² (orcid.org/ 0000-0001-5147-5047)

Berkalp TUNCA³ (orcid.org/ 0000-0002-6501-9963)

¹Eskişehir Osmangazi Üniversitesi, Turizm Fakültesi, Gastronomi ve Mutfak Sanatları Bölümü, Eskişehir, Türkiye

² Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Eskişehir, Türkiye

³Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Uygulamalı İstatistik ABD Doktora Öğrencisi, Eskişehir, Türkiye

Özet

Şarap uygarlık tarihi boyunca insanların tükettiği en özel içeceklerin başında gelmiştir. Şarap bu önemini günümüzde de sürdürmektedir. Bu nedenle şarap kalitesi gerek üreticiler gerek işletmeler gerekse tüketiciler açısından dikkate alınan bir konudur. Şarap kalitesi yakın zamana kadar sadece uzman kişilerin damak tadına bağlı olarak belirlenmiştir. Ancak şarabın kalitesini duysal etkenler kadar şarabın fizyokimyasal özellikleri de belirlemektedir. Günümüzde şarapların kalite özellikleri yapay zekâ ürünü olan makine öğrenmesi algoritmaları ile de gerçekleştirilmektedir. Bu amaçla, çalışmada, lojistik regresyon, naive bayes, k-en yakın komşu, destek vektör makineleri, karar ağaçları, doğrusal diskriminant analizi algoritmalarının kırmızı şarabın sınıflandırılmasındaki başarı oranları tespit edilmeye çalışılmıştır. En başarılı algoritmanın destek vektör makineleri algoritması olduğu belirlenmiştir. Bununla birlikte algoritma başarılarının veri seti özelliklerine göre değiştiği, bu nedenle sınıflandırmalarda veri setine en uygun algoritmanın seçilmesi gerektiği de öne çıkan bir sonuç olmuştur.

Anahtar Kelimeler: Gastronomi, Şarap, Makine Öğrenmesi, Sınıflama, Algoritma, Yapay Zekâ.

Abstract

Wine has been one of the special drinks consumed by people throughout the history of civilization. The importance of wine continues to be important in modern times. For this reason, wine quality is an issue that is taken into consideration by producers, businesses, and consumers. Until recently, wine quality has been determined only by the taste buds of experts. However, the quality of wine is determined by the physiochemical properties of the wine as well as sensory factors. Nowadays, the quality characteristics of wines are determined with machine learning algorithms as well. For this purpose, the characteristics of 1599 wines obtained from the www.kaggle.com website were classified by logistic regression, naive Bayes, k-nearest neighbor, support vector machines, decision trees, and linear discriminant analysis algorithms. It was determined that the most successful algorithm was the support vector machines algorithm. However, it was also emphasized that the algorithm success varies according to the characteristics of the data set, so the most appropriate algorithm should be selected for the data set in classification.

Keywords: Gastronomy, Wine, Machine Learning, Classification, Algorithm, Artificial Intelligence.

Giriş

Dünyanın en eski törensel içeceklerinden biri olan şarap Ön Çağ Mezopotamya uygarlıklarına kadar uzanan bir geçmişe ve üne sahiptir. İlk üretilmeye başlandığı tarihlerde üretiminin ve dağıtımının zahmetli olması sebebiyle egzotik bir içecek olarak kabul edilmekteydi. Bu sebeple de yalnızca ayrıcalıklı bir sınıf tarafından ve genellikle de dini ayinlerde tüketilmekteydi. Bu özellikler o dönem insanlarına şarabın tanrılara layık bir içki olduğunu düşündürmekteydi. Şarap, Antik Yunan döneminde felsefesi olan bir içecek haline gelmiş ve önemini günümüze kadar sürdürmüştür. Öyle ki tarih boyunca ticaretinin yapılmasından tıbbi amaçlı

*Sorumlu yazar: beybalatimur@gmail.com

DOI: 10.33083/joghat.2023.279

kullanılmasına kadar birçok alanda hayatın içinde kendisine yer bulmuştur. Bu öneminden dolayı da kalitesi ve kalitesinin belirlenmesi her zaman önem atfedilen bir konu olmuştur (Standage, 2005).

Temel olarak, kalitesinin belirlenmesinde şarabın koku ve lezzet niteliklerinin büyük bir role sahip olduğunu söylemek mümkündür (Waterhouse vd., 2016). Çünkü şarabın kimyasal yapısı, lezzetini, kokusunu, rengini ve diğer özelliklerini belirlemektedir (Sousa vd., 2014; Waterhouse vd., 2016) Kimyasal bileşim, üzüm türü, çevresel koşullar, fermantasyon sırasında mevcut olan mikrobiyal suşlar ve bağcılık uygulamalarından etkilenmektedir (Cortez vd., 2009). Yani her kimyasal bileşenin şarabın kalitesi üzerinde bir etkisi vardır demek yerinde olacaktır. Örneğin, uçucu bileşikler şaraba kokusunu, fenolik bileşikler ise lezzetini vermektedir (Sousa vd., 2014; Waterhouse vd., 2016). Tarih boyunca şarap kalitesinin üretimin sonunda şarap tadım uzmanları tarafından denenerek belirlendiği bilinmektedir. İstenilen kaliteye ulaşılamadığında çeşitli prosedürlerin en baştan uygulanması gerekmektedir. Bu durum üreticiler açısından ciddi bir maliyet anlamına gelmektedir (Dahal vd., 2021). Ayrıca, şarap tadımının insanlar tarafından gerçekleştirilmesi birtakım sorunları da beraberinde getirmektedir. Çünkü her birey hem fizyolojik (örneğin, belirli bir tat uyarana karşı duyarlılık) hem de psikolojik süreçler nedeniyle benzersiz bir şarap tadım deneyimine sahiptir. Bununla birlikte bireyin kendine özgü tecrübeleri de şarap tadımındaki algılarını etkileyebilmektedir. Bu sorunu aşmak için şarap uzmanları makul bir geçerlilik ve güvenilirlik derecesi konusunda fikir birliğine varmış olsalar da (Tiwarei vd., 2022) şarap kalitesinin belirlenmesinde insan etkeninden daha fazlasına ihtiyaç bulunduğunu söylemek mümkündür. Bu nedenle günümüzde, şarap kalitesinin belirlenmesinde pH, alkol içeriği, toplam sülfür ve antosiyanin seviyelerinin belirlenmesi için fizyokimyasal laboratuvar testleri kullanılmaktadır (Cortez vd., 2009).

Laboratuvar testlerinin sonuçlarına göre şarapların kalite sınıflandırılmaları “Makine Öğrenmesi” (MÖ) ile gerçekleştirilmektedir (Bhardwaj vd., 2022). 1930’lu yıllardan itibaren ortaya çıkan, yapay zekâ ürünü olan MÖ algoritmaları, zaman ve maliyetten tasarruf edilmesini sağladığı gibi üretim sırasında kullanılan farklı kimyasalların miktarı, sıcaklık ve üretilen şarabın kalitesi gibi çeşitli parametreleri içeren çok sayıda verinin toplanmasına da yardımcı olmaktadır. Çünkü MÖ, farklı parametrelerin kullanılmasıyla şarap kalitesinin ayarlanmasını ve farklı tatlarda ürün çıkarılmasını da sağlayabilmektedir. Bu nedenle, şarap kalitesini belirleyen temel parametrelerin analizi çok önemlidir. İnsani çabalara ek olarak, MÖ şarap kalitesini kontrol eden en önemli parametreleri belirlemek için bir alternatif olabilecektir (Dahal vd., 2021).

Son yıllarda çeşitli veri tabanlarından alınan şaraplara ait veriler kullanılarak şarap kalitesini belirlemeye yönelik bilimsel çalışmaların sayısı artmakta ve bu çalışmalarda farklı algoritmalarından yararlanılmaktadır. Bazı araştırmalarda en etkili yöntemin rastgele orman algoritmaları olduğu bildirilirken (Athanasiadis ve Ioannides 2021; Dahal vd., 2021; Er ve Atasoy 2016), destek vektör makinesi ve (Gupta, 2018) Wasserstein Üretici Çekişmeli Ağ algoritmalarının da (Doger ve Kurgun, 2021) iyi sonuçlar verdiğini bildiren çalışmalar bulunmaktadır. Ancak henüz en etkili yöntemin hangisi olduğuna dair yeterli veri olduğunu söylemek mümkün değildir. Bununla birlikte özellikle ulusal literatürde daha çok çalışmaya ihtiyaç bulunmaktadır.

Bu bilgiler ışığında, çalışmada aşağıdaki araştırma sorularına cevap aranması hedeflenmiştir:

- Şarap kalitesinin belirlenmesinde MÖ algoritmaları kullanılabilir mi?
- Şarap kalitesinin belirlenmesinde en etkili algoritma hangisi olabilecektir?

Kavramsal Çerçeve

Şarap Sektöründe Kalite ve Teknoloji

Yüzlerce yıllık geleneğin bir sonucu olarak, 1960’larda başta Fransa, İtalya, İspanya, Almanya ve Portekiz olmak üzere Avrupalı üreticiler dünya şarap üretiminin hacimce %63’ünü oluşturarak şarap endüstrisine hâkim olmuşlardır. Sektör, yerel üretimin büyük bir kısmını satın alan, büyük ve istikrarlı yerel pazarlara dayanmaktaydı. Bu dönemde kişi başına şarap tüketimi Fransa’da 124 litreye, İtalya’da 108 litreye ulaşarak dünya ortalaması olan 7,2 litrenin oldukça üzerinde gerçekleşmiştir. Şarabın küreselleşmesi hâlâ sürmekteydi ve dünya genelinde şarap üretiminin yalnızca %11’i ihraç edilmekteydi. Fransa, İtalya, Portekiz ve İspanya toplam küresel pazarın neredeyse %40’ını oluşturmaktaydı. Aynı dönemde ABD, Avustralya ve Şili gibi “Yeni Dünya” ülkelerinin şarap üretimi payı sırasıyla %2,9, %0,7 ve %1,7 olarak kaydedilmiştir. Tek büyük üretici, dünya şarap üretiminin hacimce %7,4’üne, toplam dünya tüketiminin %8’ine tekabül eden önemli bir iç pazar ve 83 litre gibi yüksek kişi başına tüketim ile Arjantin’dir. Bu tarihten itibaren, Fransa, İtalya ve İspanya’da yerel şarap tüketiminde istikrarlı bir düşüş gerçekleşmiş, 1970’lerin ortalarında hızlanan talepte bir yavaşlama ve kişi başına şarap tüketiminde sırasıyla %50 ve %59 oranlarında kümülatif bir düşüş gözlemlenmiştir. Yurt içi pazarlar incelendiğinde ise Avustralya ve ABD’de iniş ve çıkışlar yaşanırken, Şili ve Arjantin’de “Eski

Dünya” ülkelerindekiyle benzer şekilde %45’e varan sürekli bir düşüş kaydedilmiştir (Morrison ve Rabellotti, 2017). Üretim yapmayan ülkelerde ise, 1970’lerin sonundan itibaren talebin istikrarlı bir şekilde arttığını söylemek mümkündür. Bu tarihten itibaren şarap, Birleşik Krallık’ta ve İskandinav ülkeleri ile Hollanda’daki Kuzey Avrupalı tüketiciler arasında giderek daha popüler bir içecek haline gelmiştir (Anderson ve Nelgen, 2011). 1980’lerde Fransa ihracatta güç kaybederken İspanya ve İtalya yerlerini korumayı başarmışlardır. 1990’ların sonunda ise değer ve hacim olarak en büyük ikinci pazar olan ABD pazarında, Avustralya İspanya’yı geride bırakmış, İspanya da 2008’de Şili’yi geçerek ABD’nin en büyük dördüncü ihracatçısı olmuştur (Morrison ve Rabellotti, 2017).

Şarap sektörü, sadece şarabın üretildiği bölge veya ülkelerde değil, dünya ekonomisinde de büyük önem taşımaktadır. 2019 yılında dünya çapında 90.086 şirkette yaklaşık 839.908 kişinin istihdam edildiği ve küresel pazarın 315 milyar dolar olduğu tahmin edilmiştir (Ibisworld, 2021). 2023 yılı itibarıyla şarap pazarındaki gelirin 333 milyar dolar seviyesinde olduğu bilinmektedir. Küresel şarap pazarından en yüksek geliri 56,6 milyar dolar ile Amerika Birleşik Devletleri elde etmekteyken 2023 yılı sonunda şarap sektöründen kişi başı gelirin 43,35 dolar olarak kaydedileceği tahmin edilmektedir. Yine bu yıl içerisinde kişi başına ortalama şarap tüketiminin 3,14 litre olması beklenmektedir. Şarap pazarındaki 2027 yılı tahminlerinde büyüme beklentisi ise %5,5’e yükselmiştir. Yine 2027 yılında şarap pazarındaki tüketim harcamalarının %51’inin ve tüketim hacminin %26’sının ev dışı tüketimde gerçekleşmesi ve küresel şarap üretim hacminin 26 milyar litre civarında olması beklenmektedir (Statista, 2023).

Bu bilgiler ışığında, dünya genelinde şarap sektöründeki pazarın fazlasıyla değişken olduğunu ve bu pazarda ciddi bir rekabetin söz konusu olduğunu söylemek mümkündür. Pazarda yaşanan dalgalanmalar, şarap kalitesinin daha da ön plana çıkmasına sebep olmuştur. Şarap kalitesinin odağı üreticileri de aşır tüketicilere ve pazarda algılanan değere kaymıştır (Pretorius vd., 2006). Ayrıca, belirli bir şarabın itibarını oluşturma kapasitesi, görece deneyimsiz tüketicilerin büyük ve artan bir paya sahip olduğu bir pazarda önemli bir rekabet avantajı haline gelmiştir. Şarap uzmanları ve rehberleri tarafından sağlanan kalite derecelendirmeleri, potansiyel tüketicilerin algı ve davranışlarının şekillenmesinde artan bir rol oynamaktadır (Odorici ve Corrado, 2004). Bu durumun doğal bir sonucu olarak da şarap üreticileri bağıcılık, şarapçılık tekniklerini çağdaştırmak için teknoloji anlamında büyük yatırımlar yapmaya başlamışlardır. (Cusmano vd., 2010). Bu alandaki en önemli adımları en başta Avustralya ve Şili gibi Yeni Dünya ülkeleri atmıştır. Bu ülkelerde şarap sektörünü destekleyecek çeşitli araştırma merkezleri kurulmuştur. Bu merkezlerde çalışan uzmanların katkılarıyla teknolojik gelişmeler yakından takip edilerek gelişmelere ayak uydurulmaktadır. Fransa, İtalya ve ABD’de üniversitelerdeki ve araştırma enstitülerindeki araştırmacılar, yerel endüstriler için uluslararası bilimsel bilgiye önemli erişim sağlamaktadır. Sektör ve üniversiteler arasındaki bu iş birliği şaraplık üzümlerin üretimi, üretim bölgelerinin özelliklerinin belirlenmesi, şarap üretiminin daha iyi hale getirilmesi, pazarlanması, markalaşma ve kalitesinin belirlenmesi gibi konularda teknolojinin kullanımının daha da artmasının yolunu açmıştır (Morrison ve Rabellotti, 2017).

Bu çalışmaların önemli konularından biri şarap kalitesinin belirlenmesinde makine öğrenmesinden yararlanılmasıdır. Çünkü şarap kalitesinin bu yöntemle değerlendirilmesi sertifikasyon için kullanılabilir. Bu tür bir kalite sertifikası, pazarda şarap kalitesiyle ilgili bir güvence sağlayabilecektir. Şarap, yoğunluk, pH değeri, alkol ve diğer asitler gibi özelliklere sahiptir. İki tür test şarap kalitesini değerlendirebilmektedir. Bunlardan biri fizikokimyasal testler iken diğeri ise duyuşal testtir. Laboratuvar testleri fizikokimyasal özellikleri insan etkenine bağlı kalmaksızın belirleyebilmektedir, ancak duyuşal test için tadım uzmanları gerekmektedir. Ayrıca, fizikokimyasal ve duyuşal analiz arasındaki ilişkiler karmaşık olduğundan ve hala tam olarak anlaşılmadığından şarap kalitesinin sadece tadım uzmanlarına bağlı olarak değerlendirilmesi istenilen sonuçları vermeyebilir. Bu nedenle, son yıllarda, şarap kalitesinin belirlenmesinde makine öğrenmesi yöntemi giderek yaygınlaşmaktadır (Gupta, 2018).

Yapay Zekâ ve Makine Öğrenmesi

Zekâ, yaygın olarak, karmaşık problemleri çözmek için bilgi çıkarımı yapma veya bilgi birikimine sahip olma yeteneği olarak düşünülmektedir. Yapay zekâ ise bilgi toplayabilen, iletişim kurabilen, güncellenebilen ve algılayabilen akıllı makinelerin ve yazılımların incelenmesi ve geliştirilmesidir. Yapay zekâ sistemlerinin amacı insan zekâsına en yakın tasarımı oluşturmaktır. Yani makineler veya yazılım tarafından sergilenen zekâ olarak da düşünülebilir. Yapay zekâ, kalıcı, tutarlı, ucuz, çoğaltma ve yayma kolaylığına sahip olması, belgelendirilebilmesi ve belirli görevleri insandan daha hızlı yapabilmesi sayesinde doğal zekaya göre daha avantajlı bulunmaktadır (Kavuncu, 2018; Pannu, 2015).

Yapay zekânın bir alanı olan makine öğrenmesi veriden karmaşık örüntünün tespit edilmesi ve akılcı karar verme için istatistik ve bilgisayarın hesaplama gücünden faydalanmaktadır (Kaya ve Yıldız, 2014; Michalski vd., 1983; Michie vd., 1994; Nguyen ve Armitage, 2008; Sebastiani, 2002). Makine öğrenmesi, bir problemi o probleme ait veriye göre modelleyen bilgisayar algoritmalarının genel adıdır. Mevcut veri seti ve kullanılan algoritma ile oluşturulan model, en yüksek performansı vermek üzere kurulmaktadır. Bu sebeple pek çok makine öğrenmesi yöntemi geliştirilmiş olup bunlardan bazıları; k-en yakın komşu algoritması, basit (naive) Bayes sınıflandırıcı, karar ağaçları, lojistik regresyon analizi, k-ortalamlar algoritması, destek vektör makinaları ve yapay sinir ağlarıdır. Bu yaklaşımların bir kısmı tahmin ve kestirim, bir kısmı kümeleme ve bir kısmı da sınıflandırma yapabilme yeteneğine sahiptir (Atalay ve Çelik, 2017).

Makine öğrenmesi, veriden beslenen bir sisteme açık bir şekilde programlanmadan insana benzer bir şekilde otomatik olarak deneyimlerden öğrenebilme yeteneği kazandıran etkili bir yapay zekâ uygulaması olması nedeniyle pek çok alanda başarılı bir şekilde kullanılmaktadır. Ama özellikle belirsizliğin daha çok olduğu özel alanlardaki sorunların çözümünde tercih edilmektedir. Örneğin benzer belirtileri gösteren ayırt edilmesi güç hastalıkların birbirinden ayrılması, borsada hisse senetlerinin değişken hareketlerinin tahmini, istenmeyen e-postaların filtrelenmesi ve müşteri deneyimlerinden geleceğe dönük seyahat tercihlerinin tahmin edilmesi gibi pek çok konuda uygulanmaktadır (Keleş, Keleş ve Keleş 2020).

Bu yöntemlerde öğrenme stratejileri; denetimli, denetimsiz ve pekiştirmeli (takviyeli) olmak üzere üç grupta incelenmektedir. Denetimli öğrenmede oluşturulan model ile, bir grup girdi değerine karşılık onlara ait hedef değerleri verilerek aralarındaki ilişkiyi öğrenmesi ve hedef değerlere en yakın çıktılarını üretmesi amaçlanmaktadır. Elde edilen en iyi model, yeni girdi değerleri için en yakın çıktıyı da verebilecektir. Denetimsiz öğrenmede ise hedef değerleri olmadan sadece girdi değerleri arasındaki ilişki ortaya çıkarılmaya çalışılır. Bu ilişki(ler) yardımı ile birbirine yakın değerler gruplandırılır yani kümeleme yapılır. Yeni bir girdi bu kümelere hangisiyle ilişkili ise o kümeyle ait olacaktır. Pekiştirmeli (takviyeli) öğrenme yönteminde, hedef çıktıyı vermek için bir danışman yerine, elde edilen çıkışın verilen girişe karşılık iyi ya da kötü olarak değerlendirilen bir kriter kullanılmaktadır (Atalay ve Çelik, 2017). Makine öğrenmesinde kullanılan yöntemler şu şekilde sınıflandırılabilir:

- **Denetimli Öğrenme (Supervised Learning):** Denetimli öğrenmede amaç, bir dizi girdi değişkenine dayanarak çıktı değerini tahmin etmektir (Hastie vd., 2009). Bu tür yaklaşımda öğrenen sisteme öğrenmesi istenen olay ile ilgili örnekler girdi/çıktı seti olarak verilir. Sistemin görevi girdileri belirlediği çıktılara haritalamaktır. Böylece olayın girdileri ile çıktılarının arasındaki ilişkiler öğrenilmektedir (Öztemel, 2012). Denetimli makine öğrenme yaklaşımı tahmin ve kestirim problemlerine odaklanmaktadır. Denetimli makine öğrenmesi regresyon ve sınıflandırma olarak iki alt grupta ele alınmaktadır (Çağlayan Akay, 2018; Pulat ve Kocakoç Deveci, 2021).
- **Denetimsiz Öğrenme (Unsupervised Learning):** Denetimsiz öğrenmede çıktı değeri yoktur ve amaç bir dizi girdi değişkeni arasındaki ilişkileri ve örüntüleri tanımlamaktır (Hastie vd., 2009). Bu tür stratejide sisteme sadece girdi değerleri gösterilir ve sadece girdiler üzerinde öğrenme işlemi gerçekleştirilmeye çalışılır. Örneklerdeki parametreler arasındaki ilişkileri sistemin kendi kendisine öğrenerek ilişki ağını ortaya koyması beklenir (Öztemel, 2012). Denetimsiz öğrenme stratejisi kümeleme, olasılık yoğunluk tahmini, öznitelikler arasındaki ilişkilerin keşfedilmesi ve boyut indirgeme gibi amaçlar için kullanılmaktadır (Kutlugün, 2017; Pulat ve Kocakoç Deveci, 2021).
- **Takviyeli Öğrenme (Reinforcement Learning):** Bu yaklaşımda öğrenen sisteme her girdi seti için üretilmesi gereken çıktı setini sisteme vermek yerine sistemin kendisine gösterilen girdilere karşılık çıktıyı üretmesi beklenir ve ürettiği çıktının doğru veya yanlış olduğuna dair bir sinyal üretir. Sistem üretilen sinyali göz önüne alarak öğrenme sürecini devam ettirir (Öztemel, 2012; Pulat ve Kocakoç Deveci, 2021).

Makine öğrenmesinde kullanılan bazı yöntemler ise aşağıdaki gibi ifade edilmiştir:

Lojistik Regresyon

Lojistik regresyon, ikili sınıfları öngörmek için istatistiksel bir yöntemdir. Lojistik regresyon, yalnızca iki değere sahip olabilen bir sonucun olasılığını öngörür. Tahmin, bir veya birkaç öngörücünün (sayısal ve kategorik) kullanımına dayanır. Doğrusal regresyon evet/hayır, var/yok gibi binary (ikili) sistemde ifade edilebilecek değerler için uygun değildir. Çünkü, 0 ve 1 aralığının dışında değer tahmin edebilir. Lojistik

regresyon, 0 ile 1 arasındaki değerlerle sınırlı lojistik eğrisi üretir (Kazan ve Karakoca, 2019; Kleinbaum ve Klein, 2010). Bir $a+bx$ denklemi için olayın gerçekleşme olasılığı Denklem 1'deki gibidir.

Denklem 1. $a+bx$ Denklemi İçin Olayın Gerçekleşme Olasılığı

$$p = \frac{e^{a+bx}}{1 + e^{a+bx}}$$

Olayın gerçekleşme olasılığı ise $1-p$ olmak üzere logit fonksiyonu Denklem 2'de verilmiştir.

Denklem 2. Logit Fonksiyonu

$$\text{logit}(p) = \ln \frac{p}{1-p}$$

Naive Bayes

Naive Bayes (NB) sınıflandırma Bayes teoremine dayanır ve hedef sınıfı için verilen değerlerin gerçekleşme olasılığının ne olduğunu bildirir. Denklem 3'teki gösterilen teoreme dayanır (Kazan ve Karakoca 2019; Wood vd., 2019).

Denklem 3. Bayes Teoremi Denklemi

$$P(c|x) = \frac{P(c|x)P(x)}{P(x)}$$

Formüle göre, $P(c|x)$, c olayının verilen x olayına göre olma olasılığıdır. $P(x|c)$ ise x olayı gerçekleştiğinde c olayının gerçekleşme olasılığıdır. $P(x)$ ve $P(c)$ ise x ve c olaylarının ösel olasılıklarıdır.

K-En Yakın Komşu

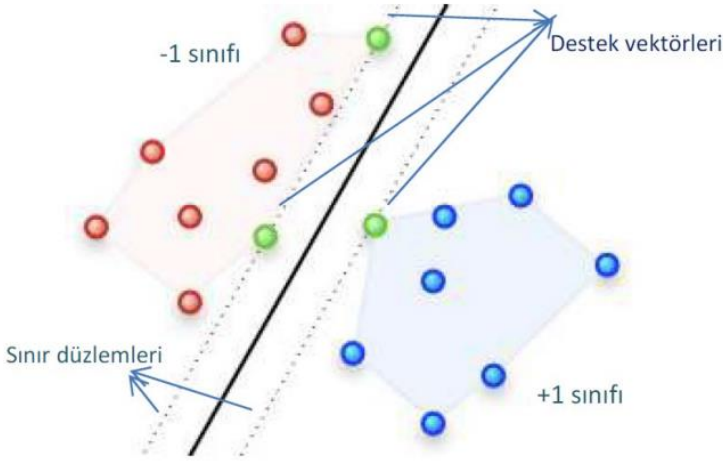
Sınıfı belirlenmek istenen bir noktanın, daha önceden sınıflanmış olan noktalardan, belirlenen K sayısının en yakın noktaya göre sınıfının tespit edilmesini sağlayan bir modeldir. En yakın noktalar hesaplanırken genelde öklit uzaklığına bakılır. İdeal K değerinin seçimi üzerinde çalışılan veriye bağlı olarak değişiklik gösterir. Büyük K değerleri sınıflamadaki gürültü etkisini azaltırken, sınıflar arasındaki sınırların ayırımı azaltır.

Karar Ağaçları

Karar ağacı hem regresyon hem de sınıflandırma problemlerinde kullanılabilen denetimli bir öğrenme algoritması türüdür. Hem kategorik hem de sürekli giriş ve çıkış değişkenleri için çalışır. Karar Ağacı olası tüm eylem seçeneklerini, bu eylem seçeneklerine etkisi olabilecek tüm olası faktörleri ve tüm bu faktörlere dayanan her bir olası sonucu, verilere bağlı olarak değerlendirir, çizgi, kare, daire gibi geometrik semboller kullanımı yoluyla karar vericiye problemi anlamada kolaylık sağlayan grafiksel bir teknik olarak tanımlanabilir. Karar Ağacı, grafik gösterimi ile problemin tüm yönlerini ayrıntılı olarak ortaya koymaktadır. Herhangi bir karar problemi için kullanılabilen Karar Ağacı tekniği özellikle birden fazla kararın ardışık olarak verilmesini gerektiren karar problemlerinin gösteriminde çok kullanışlıdır (Albright, Winston ve Zappe 2011; Seveli 2019).

Destek Vektör Makineleri

Destek Vektör Makineleri (DVM) öğrenme, sınıflandırma, kümeleme, yoğunluk tahmini ve son olarak da veriden regresyon kuralları üretmek için kullanılan eğitim algoritmasıdır. DVM iki sınıflı ve çok sınıflı sınıflandırma probleminin çözümü için kullanılabilir. DVM veriyi sınıflandırırken sınıfların birbirlerine en yakın örneklerini bularak bu örneklerin (iki sınıfı ayıracak olan) ayırıcı yüzeye dik uzaklıklarını maksimize etmeyi amaçlar. Ayırıcı yüzeyin, veri kümesi üzerindeki başarısı değişmeden birçok farklı alternatifi olabilir. DVM sayesinde ayırıcı yüzey her iki sınıfa da aynı mesafede ve maksimum uzaklıktadır (Gunn, 1998; Kazan ve Karakoca, 2019). Şekil 1'de DVM için iki sınıflı problem örneği gösterilmiştir.

Şekil 1. İki Sınıflı Bir Problem İçin DVM Örneği**Doğrusal Diskriminant Analizi**

Doğrusal Diskriminant Analizi (DDA), veri boyutunu indirgemede sıklıkla kullanılmaktadır (Balakrishnama ve Ganapathiraju, 1998). DDA, sınıf içi frekansların eşit olmadığı ve performanslarının rasgele oluşturulmuş test verileri üzerinde incelendiği durumu ele alır. Bu yöntem, herhangi bir belirli veri setinde sınıflar arası varyansın sınıf içi varyansa oranını maksimuma çıkarır, böylece maksimum ayrılabilirliği sağlamaktadır (Badem, 2019; Ming Li ve Baozong Yuan, 2004).

Yöntem

Bu çalışmanın amacı kırmızı şarap kalitesini sınıflarken, makine öğrenme algoritmalarının başarılarını karşılaştırmaktır. Bu amaçla, Cortez vd.'nin (2009) çalışmasında kullandıkları ve <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009> adresinde bulunan 1599 adet kırmızı şaraba ilişkin fizyokimyasal ve duyuşal çıktı değişkenleri ele alınmıştır (Kaggle, 2019). Verilerin analizi Python programında derlenmiştir. Cortez vd. (2009) mevcut veri seti ile çoklu regresyon, yapay sinir ağları ve destek vektör makineleri algoritmalarını karşılaştırırken bu çalışmada lojistik regresyon, naive bayes, k-en yakın komşu, karar ağacı, destek vektör makineleri, doğrusal destek vektör makineleri ve doğrusal diskriminant analizi teknikleri karşılaştırılmıştır.

Tablo 1. Değişkenler ve Tanımlarına İlişkin Bilgiler

Değişken	Açıklama
Sabit asit	Şarapla ilgili, sabit veya uçucu olmayan asitlerin çoğu (kolayca buharlaşmaz)
Uçucu asitlik	Şarapta çok yüksek seviyelerde tatsız, sirke tadına yol açabilen asetik asit miktarı
Sitrik asit	Küçük miktarlarda bulunan sitrik asit, şaraplara 'tazelik' ve lezzet katabilir
Artık şeker	Fermentasyon durduktan sonra kalan şeker miktarı
Klorürler	Şaraptaki tuz miktarı
Serbest kükürt dioksit	Serbest SO ₂ formu, moleküler SO ₂ (çözünmüş bir gaz olarak) ve bisülfid iyonu arasında denge halinde bulunur; mikrobiyal büyümeyi ve şarabın oksidasyonunu önler.
Toplam kükürt dioksit	SO ₂ 'nin serbest ve bağlı formlarının miktarı; düşük konsantrasyonlarda, SO ₂ çoğunlukla şarapta saptanamaz, ancak 50 ppm'in üzerindeki serbest SO ₂ konsantrasyonlarında, SO ₂ şarabın burnunda ve tadında belirginleşir
Yoğunluk	Yüzde alkol ve şeker içeriğine bağlı olarak şarabın yoğunluğu suyunkine yakındır
pH	Bir şarabın ne kadar asidik veya bazik olduğunu 0 (çok asidik) ile 14 (çok bazik) arasında bir ölçekte tanımlar; çoğu şarap pH ölçeğinde 3-4 arasındadır
Sülfatlar	Kükürt dioksit gazı (SO ₂) seviyelerine katkıda bulunabilen, antimikrobiyal ve antioksidan görevi gören bir şarap katkı maddesi.
Alkol	Şarabın yüzde alkol içeriği
Kalite puanları	Çıkış değişkeni (duyuşal verilere göre): kalite (0 ile 10 arasında puan). Var olan kalite puanlar; 3, 4, 5, 6, 7, 8

Kaynak: (Gupta, 2018)

Şarap kalitesinin belirlenmesinde sabit asit, uçucu asitlik, sitrik asit, artık şeker, klorürler, serbest kükürt dioksit, toplam kükürt dioksit, yoğunluk ve pH değeri, sülfat, alkol içeriği ve duyu kalite puanları değişken olarak belirlenmiştir (Gupta, 2018). Veri setinin temsili görüntüsü Tablo 2’de sunulmuştur.

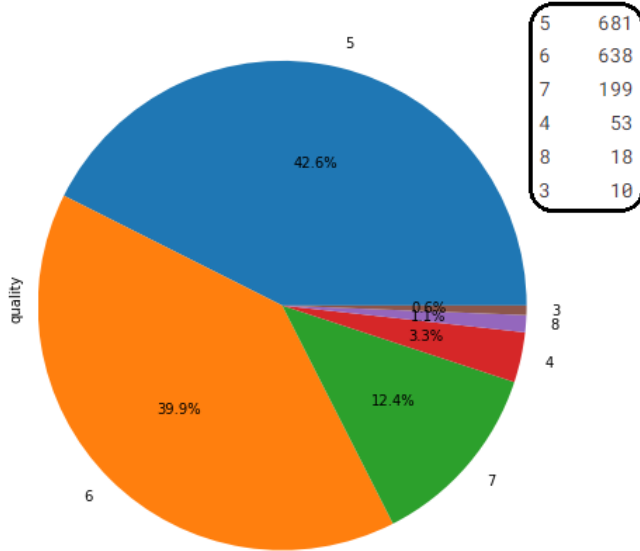
Tablo 2. Veri Setinin Temsili Görüntüsü

Sabit asit	Uçucu asitlik	Sitrik asit	Artık şeker	Klorürler	Serbest kükürt dioksit	Toplam kükürt dioksit	Yoğunluk	pH	Sülfatlar	Alkol	Kalite puanları
7.4	0.70	0.00	1.9	0.08	11.0	34.0	1.00	3.51	0.56	9.4	5
7.8	0.88	0.00	2.6	0.10	25.0	67.0	1.00	3.20	0.68	9.8	5
7.8	0.76	0.04	2.3	0.09	15.0	54.0	1.00	3.26	0.65	9.8	5
11.2	0.28	0.56	1.9	0.07	17.0	60.0	1.00	3.16	0.58	9.8	6
7.4	0.70	0.00	1.9	0.08	11.0	34.0	1.00	3.51	0.56	9.4	5

Bulgular

Veriler analiz edilmeden önce veri seti örnekleri (Cortez et al. 2009) incelenmiştir. İnceleme sonucunda puanların sırasıyla en yüksekte en düşüğe doğru 5 (%42,6), 6 (%39,9), 7 (%12,4), 4 (%3,3), 8 (%1,1) ve 3 (%0,6) olduğu görülmüştür (Şekil 2). Bu sıralama, veri setindeki örneklerde sınıf dağılımının dengesiz olduğuna işaret etmektedir. Bu durum makine öğrenmesi modelleri için istenmeyen bir durumdur (Aydın, 2021). Bu nedenle “stratify” (tabakalaştırma) özelliği kullanılarak az örneğe sahip sınıflar çoğaltılmış ve sınıflardaki örnek sayıları eşitlenmiştir.

Şekil 2. Veri Setindeki Kalite Puanlarının Dağılımı



Veri setindeki diğer değişkenlere ait ortalamalar incelendiğinde sabit asit ortalamasının 8.35, uçucu asitlik ortalamasının 0.53, sitrik asit ortalamasının 0.27, artık şeker ortalamasının 2.54, klorür ortalamasının 0.09, serbest kükürt dioksit ortalamasının 15.87, toplam kükürt dioksit ortalamasının 46.47, yoğunluk ortalamasının 1, pH ortalamasının 3.31, sülfat ortalamasının 0.66, alkol ortalamasının 10.43 ve kalite puan ortalamasının 5.64 olduğu görülmektedir (Tablo. 3).

Tablo 3. Verilerin Dağılımına İlişkin Tanımlayıcı İstatistikler

	Frekans	Ortalama	Std. Sapma	Min	25%	50%	75%	Max
Sabit asit	1599	8.35	1.74	4.60	7.10	7.90	9.20	15.90
Uçucu asitlik	1599	0.53	0.18	0.12	0.39	0.52	0.64	1.58
Sitrik asit	1599	0.27	0.19	0.00	0.09	0.26	0.42	1.00
Artık şeker	1599	2.54	1.44	0.90	1.90	2.20	2.60	15.50
Klorürler	1599	0.09	0.05	0.01	0.07	0.08	0.09	0.61
Serbest kükürt dioksit	1599	15.87	10.46	1.00	7.00	14.00	21.00	72.00
Toplam kükürt dioksit	1599	46.47	32.89	6.00	22.00	38.00	62.00	289.00
Yoğunluk	1599	1.00	0.01	0.99	0.99	1.00	1.00	1.00
pH	1599	3.31	0.15	2.74	3.21	3.31	3.40	4.01
Sülfatlar	1599	0.66	0.17	0.33	0.55	0.62	0.73	2.00
Alkol	1599	10.43	1.07	8.40	9.50	10.20	11.10	14.90
Kalite puanları	1599	5.64	0.81	3.00	5.00	6.00	6.00	8.00

Değişkenlerin kalite puanı bazında tanımlayıcı istatistikleri incelendiğinde ortalama alkol değerinin kaliteyle doğru orantılı olarak az da olsa arttığı görülmüştür (Tablo 4). Ortalama uçucu asitlik değeri kalite seviyesi arttıkça azalmaktadır. Ortalama sülfatlar değeri kalite seviyesi arttıkça azalmaktadır.

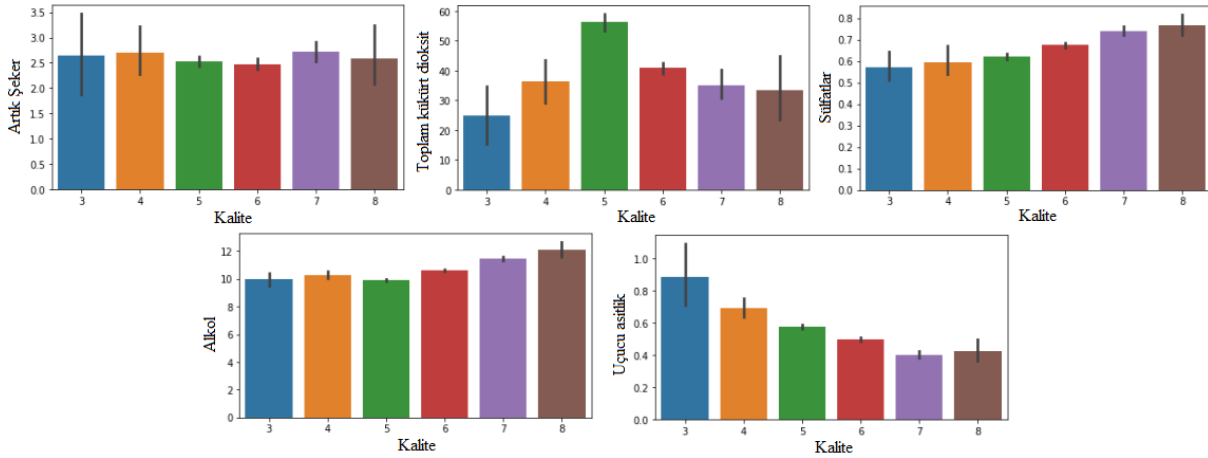
Tablo 4. Kalite Puanı Bazında Değişkenlere Ait Veriler

	Değişkenler	Frekans	Ortalama	Standart Sapma
Kalite seviyesi: 3,4	Artık şeker	63	2.68	1.72
	Toplam kükürt dioksit	63	34.44	26.40
	Sülfatlar	63	0.59	0.22
	Alkol	63	10.22	0.92
	Uçucu asitlik	63	0.72	0.25
	Kalite	63	3.84	0.37
Kalite seviyesi: 5,6	Artık şeker	1319	2.50	1.40
	Toplam kükürt dioksit	1319	48.95	32.71
	Sülfatlar	1319	0.65	0.17
	Alkol	1319	10.25	0.97
	Uçucu asitlik	1319	0.54	0.17
	Kalite	1319	5.48	0.50
	Artık şeker	217	2.71	1.36

Kalite seviyesi: 7,8	Toplam kükürt dioksit	217	34.89	32.57
	Sülfatlar	217	0.74	0.13
	Alkol	217	11.52	1.00
	Uçucu asitlik	217	0.41	0.14
	Kalite	217	7.08	0.28

Ortalama artık şeker ve toplam kükürt dioksit değerlerinin kalite seviyesine göre farklılıkları olmakla beraber kalite seviyesiyle artan veya azalan bir ilişkiye sahip değillerdir (Şekil 3).

Şekil 3. Kalite Puanı Bazında Bazı Değişkenlerin Bar Grafikleri



Veri setindeki değişkenler arası korelasyonlar incelendiğinde sitrik asit ile sabit asit arasındaki, yoğunluk ile sabit asit arasındaki, toplam kükürt dioksit ile serbest kükürt dioksit arasındaki korelasyon katsayıları pozitif ve orta-güçlü arasında olduğu görülmektedir. pH ile sabit asit arasındaki, pH ile sabit asit arasındaki, sitrik asit ile uçucu arasındaki, alkol ile yoğunluk arasındaki korelasyonlar ise negatif yönde ve orta derecededir.

Veri seti %70 train, %30 test olarak bölünmüştür. Buna göre hesaplanan çapraz doğrulama başarı oranları Tablo 5'te verilmiştir.

Tablo 5. 6'lı Çapraz Doğrulama Sonuçlarına İlişkin Başarı Oranları

Makine Öğrenmesi Algoritması	Başarı Yüzdesi
Lojistik Regresyon	0.58
Naive Bayes	0.56
K-En Yakın Komşu	0.58
Karar Ağacı	0.63
Destek Vektör Makineleri	0.62
Doğrusal Destek Vektör Makineleri	0.57
Doğrusal Diskriminant Analizi	0.59

Analiz sonuçlarına göre, en iyi sınıflamayı 0.63 ile Karar Ağacı algoritması, en kötü sınıflamayı ise 0.56 ile Naive Bayes algoritması yapmıştır. Ancak algoritmalar arasındaki başarı yüzdelerinin birbirine çok yakın olduğunu da vurgulamak gerekmektedir. Başarı oranlarının doğruluğunu tespit etmek için GridSearch ile hiperparametrelerin optimizasyonu yapılmış ve doğruluk oranı 0.64 olarak hesaplanmıştır.

Burada 6 farklı kalite için kurulan algoritmalar varsayılan olarak belirlenen hiperparametreler üzerinden uygulanmıştır. Kalite sınıfının düşürülmesinin sınıflama başarısını arttırıldığı düşünülerek sınıf sayısı 3'e düşürülmüştür (3, 4=1 (n=1319), 5,6=2 (n=217) ve 7,8=3 (n=63)).

Tablo 6. 3'lü Çapraz Doğrulama Sonuçlarına İlişkin Başarı Oranları

Makine Öğrenmesi Algoritması	Başarı Yüzdesi
Lojistik Regresyon	0.84

Naive Bayes	0.78
K-En Yakın Komşu	0.83
Karar Ağacı	0.82
Destek Vektör Makineleri	0.85
Doğrusal Destek Vektör Makineleri	0.83
Doğrusal Diskriminant Analizi	0.83

Çapraz doğrulama sonuçlarına göre, en iyi sınıflamayı 0.85 ile Destek Vektör Makineleri, en kötü sınıflamayı 0.78 ile yine Naive Bayes algoritması yapmıştır. GridSearch tekniği yardımıyla hiperparametrelerin optimizasyonu sonucunda bu bulguların doğruluk oranı 0.86 olarak hesaplanmıştır.

Tartışma ve Sonuç

Bu çalışmada, ele alınan şarap kalitesini değerlendirmede kullanılan sabit asit, uçucu asitlik, sitrik asit, artık şeker, klorürler, serbest kükürt dioksit, toplam kükürt dioksit, yoğunluk, pH, sülfatlar, alkol değişkenleri için şarap kalitesine göre sınıflandırılma yapılmıştır. Bu sınıflandırma için makine öğrenme algoritmalarından olan lojistik regresyon, naive bayes, k-en yakın komşu, karar ağacı, destek vektör makineleri, doğrusal destek vektör makineleri ve doğrusal diskriminant analizi teknikleri hakkında kısa bilgiler verilerek analizler gerçekleştirilmiştir. Bu teknikler sınıflama başarı oranlarına göre karşılaştırılmıştır.

Analizler gerçekleştirilirken öncelikle 6 sınıf kalitesine göre sınıflandırılmıştır. Bu 6 sınıf kalitesi için çapraz doğrulama sonuçlarında en iyi başarıyı karar ağacı algoritması yapmıştır. En kötü sınıflamayı ise naive bayes algoritması yapmıştır. Naive bayes algoritmasının en kötü çıkmasının sebebi, koşullu olasılık temelli olan bayes teoreminin küçük örneklerde daha iyi sonuçlar sağlamasından dolayı olabileceği düşünülmektedir.

Daha sonrasında kalite sınıfının 6 yerine 3 olduğu durumdaki sınıflama başarıları test edilmiştir. Çünkü sınıf sayısının düşürülmesinin çapraz doğrulama sonucunu olumlu oranda değiştireceği ön görülmüştür. Bu bağlamda şarap kaliteleri 3 sınıf altında toplanarak aynı şartlarda analizler tekrar gerçekleştirilmiştir. Analizlerin sonucunda ise, çapraz doğrulama sonuçlarının tahmin edildiği gibi arttığı, yani sınıflama başarı oranlarının arttığı görülmüştür. Bu sonuçlar incelendiğinde ise, çapraz doğrulama oranı en yüksek destek vektör makineleri algoritması, en düşük yine naive bayes algoritması olduğu görülmüştür. Bu çalışmada da görüldüğü gibi aslında çıktı değişkeninin kategori sayısının artırılmasının, başarı oranını düşürdüğü söylenebilir.

Ayrıca literatürdeki makine öğrenmesi algoritmalarının sınıflama başarıları karşılaştırma sonuçları incelendiğinde, yapay sinir ağları (Alpkoçak vd., 2019; Özgür ve Erdem 2012), naive bayes (Karakoyun ve Hacıbeyoğlu, 2014), k-en yakın komşu (Kaynar vd., 2018), karar ağaçları (Özlüer Başer, Yangın ve Sarıdaş 2021) algoritmalarının diğerlerine göre üstünlükleri görülmüştür. Uluslararası literatürde de çoğunlukla rastgele ormanlar algoritmasının diğerlerine üstünlük sağladığı bildirilmiştir (Athanasiadis ve Ioannides 2021; Tiwari vd., 2022). Bazı araştırmalarda da AdaBoost ve Gradient Boosting algoritmaları başarılı bulunmuştur (Bhardwaj vd., 2022; Dahal vd., 2021). Bu sonuçların ışığında, makine öğrenme algoritmalarının veri setine göre başarı oranlarının değişiklik gösterdiği görülmektedir. Dolayısıyla herhangi bir algoritmanın diğeri üzerinde belirli bir üstünlüğünden bahsedilmesi söz konusu olmayıp, bütün teknikleri uygulayarak, en iyi sınıflama sonucunu hangi algoritma veriyorsa araştırmacıların ona göre çalışmalarını sürdürmesi önerilmektedir.

Son olarak, bu çalışmada makine öğrenme algoritmaları uygulanırken hiperparametreler için farklı denemeler yapılarak uygulanmamıştır. Sadece hiperparametrelerin optimizasyonu için GridSearch tekniği uygulanmıştır. Gelecek çalışmalar için farklı parametre tahminleri başarı oranını arttırabileceği öngörülmektedir.

Kaynakça

- Albright, C. S., Winston, W. L., ve Zappe, C. J. (2011). *Data analysis ve decision making* (4. bs). South-Western CENGAGE Learning.
- Alpkoçak, A., Tocoglu, M. A., Çelikten, A., ve Aygün, İ. (2019). Türkçe metinlerde duygu analizi için farklı makine öğrenmesi yöntemlerinin karşılaştırılması. *Deu Muhendislik Fakültesi Fen ve Muhendislik*, 21(63), 719-725. <https://doi.org/10.21205/deufmd.2019216303>
- Anderson, K., ve Nelgen, S. (2011). Wine's globalization: new opportunities, new challenges. *Wine Economics Research Centre Working Papers*, 0111. <https://ideas.repec.org/p/adl/winewp/2011-01.html>

- Atalay, M., ve Çelik, 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(22), 155-172. <https://doi.org/10.20875/MAKUSOBED.309727>
- Athanasiadis, I., ve Ioannides, D. (2021). A machine learning approach using random forest and LASSO to predict wine quality. *International Journal of Sustainable Agricultural Management and Informatics*, 7(3), 232-251. <https://doi.org/10.1504/IJSAMI.2021.118129>
- Badem, H. (2019). Parkinson hastalığının ses sinyalleri üzerinden makine öğrenmesi teknikleri ile tanımlanması. *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 2, 630-637. <https://doi.org/10.28948/ngumuh.524658>
- Balakrishnama, S., ve Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 18(1998), 1-8.
- Bhardwaj, P., Tiwari, P., Olejar, K., Parr, W., ve Kulasiri, D. (2022). A machine learning application in wine quality prediction. *Machine Learning with Applications*, 8, 100261. <https://doi.org/10.1016/j.mlwa.2022.100261>
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., ve Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4), 547-553. <https://doi.org/10.1016/j.dss.2009.05.016>
- Cusmano, L., Morrison, A., ve Rabellotti, R. (2010). Catching up trajectories in the wine sector: a comparative study of Chile, Italy, and South Africa. *World Development*, 38(11), 1588-1602. <https://doi.org/10.1016/j.worlddev.2010.05.002>
- Çağlayan Akay, E. (2018). Ekonometride yeni bir ufuk: büyük veri ve makine öğrenmesi. *Sosyal Bilimler Araştırma Dergisi*, 7(2), 41-53.
- Dahal, K. R., Dahal, J. N., Banjade, H., Gaire, S., Dahal, K. R., Dahal, J. N., Banjade, H., ve Gaire, S. (2021). Prediction of wine quality using machine learning algorithms. *Open Journal of Statistics*, 11(2), 278-289. <https://doi.org/10.4236/OJS.2021.112015>
- Doger, Ş., ve Kurgun, O. A. (2021). Şarap üretiminde veri kalitesine ilişkin eksik veri sorunlarının derin öğrenme ile çözülmesi: üretici çekişmecî ağlarla bir uygulama. *International Journal of Contemporary Tourism Research*, 5(1), 99-111. <https://doi.org/10.30625/ijctr.943818>
- Er, Y., ve Atasoy, A. (2016). The classification of white wine and red wine according to their physicochemical qualities. *International Journal of Intelligent Systems and Applications in Engineering*, 4(Special Issue-1), 23-26. <https://doi.org/10.18201/IJISAE.265954>
- Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS technical report*, 14(1), 5-16.
- Gupta, Y. (2018). Selection of important features and predicting wine quality using machine learning techniques. *Procedia Computer Science*, 125, 305-312. <https://doi.org/10.1016/J.PROCS.2017.12.041>
- Hastie, T., Tibshirani, R., ve Friedman, J. (2009). *The elements of statistical learning* (2. bs). Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Ibisworld. (2021). *Global wine manufacturing -industry data, trends, stats | IBISWorld*. Global Wine Manufacturing Industry - Market Research Report. Erişim Tarihi: 15.04.2023, <https://www.ibisworld.com/global/market-research-reports/global-wine-manufacturing-industry/>
- Kaggle. (2019). *Red wine quality*. Erişim tarihi: 23.06.2023, <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>.
- Karakoyun, M., ve Hacıbeyoğlu, M. (2014). Biyomedikal veri kümeleri ile makine öğrenmesi sınıflandırma algoritmalarının istatistiksel olarak karşılaştırılması. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 16(48), 30-42.
- Kavuncu, S. K. (2018). *Makine öğrenmesi ve derin öğrenme: Nesne tanıma uygulaması* (Yüksek Lisans Tezi). Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü, Kırıkkale.

- Kaya, Ç., ve Yıldız, O. (2014). Makine öğrenmesi teknikleriyle saldırı tespiti: karşılaştırmalı analiz. *Marmara Fen Bilimleri Dergisi*, 3, 89-104.
- Kaynar, O., Arslan, H., Görmez, Y., ve Işık, Y. E. (2018). Makine öğrenmesi ve öznelik seçim yöntemleriyle saldırı tespiti. *Bilişim Teknolojileri Dergisi*, 11(2), 175-185. <https://doi.org/10.17671/gazibtd.368583>
- Kazan, S., ve Karakoca, H. (2019). Makine öğrenmesi ile ürün kategorisi sınıflandırma. *Sakarya University Journal of Computer and Information Sciences*, 2(1), 18-27. <http://saucis.sakarya.edu.tr/en/download/article-file/702149>
- Keleş, M. B., Keleş, A., ve Keleş, A. (2020). Makine öğrenmesi yöntemleri ile uçuş fiyatlarının tahmini. *Euroasia Journal of Mathematics, Engineering, Natural ve Medical Sciences*, 7(11), 72-78. <https://euroasiajournal.org/index.php/ejas/article/view/93>
- Kleinbaum, D. G., ve Klein, M. (2010). *Logistic regression* (3. bs). New York: Springer. <https://doi.org/10.1007/978-1-4419-1742-3>
- Kutlugün, M. A. (2017). *Gözetimli makine öğrenmesi yoluyla türe göre metinden ses sentezleme* (Yüksek Lisans Tezi). İstanbul Zaim Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Michalski, R. S., Carbonell, J. G., ve Mitchell, T. M. (1983). *Machine learning*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-12405-5>
- Michie, D., Spiegelhalter, D., ve Taylor, C. (1994). *Machine learning, neural and statistical classification*. Ellis Horwood Limited.
- Ming Li, ve Baozong Yuan. (2004). A novel statistical linear discriminant analysis for image matrix: two-dimensional fisherfaces. *Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004.*, 1419-1422. <https://doi.org/10.1109/ICOSP.2004.1441592>
- Morrison, A., ve Rabelotti, R. (2017). Gradual catch up and enduring leadership in the global wine industry. *Research Policy*, 46(2), 417-430. <https://doi.org/10.1016/j.respol.2016.09.007>
- Nguyen, T. T. T., ve Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys ve Tutorials*, 10(4), 56-76. <https://doi.org/10.1109/SURV.2008.080406>
- Odorici, V., ve Corrado, R. (2004). Between supply and demand: intermediaries, social networks and the construction of quality in the italian wine industry. *Journal of Management and Governance*, 8(2), 149-171. <https://doi.org/10.1023/B:MAGO.0000026542.18647.48>
- Özgür, A., ve Erdem, H. (2012). Saldırı tespit sistemlerinde kullanılan kolay erişilen makine öğrenme algoritmalarının karşılaştırılması. *Bilişim Teknolojileri Dergisi*, 5(2), 41-48.
- Özlüer Başer, B., Yangın, M., ve Sarıdaş, E. S. (2021). Makine öğrenmesi teknikleriyle diyabet hastalığının sınıflandırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 25(1), 112-120. <https://doi.org/10.19113/sdufenbed.842460>
- Öztemel, E. (2012). *Yapay sinir ağları* (3. bs). Papatya Yayıncılık Eğitim.
- Pannu, A. (2015). Artificial intelligence and its application in different areas. *International Journal of Engineering and Innovative Technology (IJEIT)*, 4(10), 79-84.
- Pretorius, I. S., Bartowsky, E. J., Bauer, F., De Barros Lopes, M., Du Toit, M., Van Rensburg, P., ve Vivier, M. A. (2006). The tailoring of designer grapevines and microbial starter strains for a market-directed and quality-focussed wine industry. İçinde Y. H. Hui (Ed.), *Handbook of Food Science, Technology, and Engineering* (4, 406-428).
- Pulat, M., ve Kocakoç Deveci, İ. (2021). Türkiye’de makine öğrenmesi ve karar ağaçları alanında yayınlanmış tezlerin bibliyometrik analizi. *Journal of Management and Economics*, 28(2), 287-308. <https://doi.org/10.18657/YONVEEK.870190>
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, 34(1), 1-47. <https://doi.org/10.1145/505282.505283>

- Sevli, O. (2019). Göğüs kanseri teşhisinde farklı makine öğrenmesi tekniklerinin performans karşılaştırması. *European Journal of Science and Technology*, 16, 176-185. <https://doi.org/10.31590/ejosat.553549>
- Sousa, E. C., Uchôa-Thomaz, A. M. A., Carioca, J. O. B., Morais, S. M. de, Lima, A. de, Martins, C. G., Alexandrino, C. D., Ferreira, P. A. T., Rodrigues, A. L. M., Rodrigues, S. P., Silva, J. do N., ve Rodrigues, L. L. (2014). Chemical composition and bioactive compounds of grape pomace (*Vitis vinifera* L.), Benitaka variety, grown in the semiarid region of Northeast Brazil. *Food Science and Technology (Campinas)*, 34(1), 135-142. <https://doi.org/10.1590/S0101-20612014000100020>
- Standage, T. (2005). *A history of the world in 6 glasses*. Walker ve Company.
- Statista. (2023). *Wine-worldwide*. Erişim Tarihi: 15.04.2023, <https://www.statista.com/outlook/cmo/alcoholic-drinks/wine/worldwide>
- Tiwari, P., Bhardwaj, P., Somin, S., Parr, W. V., Harrison, R., ve Kulasiri, D. (2022). Understanding quality of pinot noir wine: can modelling and machine learning pave the way? *Foods*, 11(19). <https://doi.org/10.3390/foods11193072>
- Waterhouse, A. L., Sacks, G. L., ve Jeffery, D. W. (2016). Grape genetics, chemistry, and breeding. İçinde *Understanding Wine Chemistry* (s. 2-5). John Wiley ve Sons. <https://doi.org/10.1002/9781118730720>
- Wood, A., Shpilrain, V., Najarian, K., ve Kahrobaei, D. (2019). Private naive bayes classification of personal biomedical data: Application in cancer data analysis. *Computers in Biology and Medicine*, 105, 144-150. <https://doi.org/10.1016/j.combiomed.2018.11.018>