



Makine Öğrenmesi Algoritmalarını Kullanarak 1990-2023 Yılları Arası Veriler ile Türkiye'de Et Tüketim Talebinin Tahmini Üzerine Kapsamlı Bir Çalışma

Hasan Ali AKYÜREK¹

Hasan İbrahim KOZAN²

Öz

Çok eski zamanlardan beri et, insan diyetinin temel bir parçası olmuştur ve birçok medeniyette temel kaynaklar arasında halen yer almaktadır. Farklı ülkeler ve kültürler arasında tüketilen et miktarı ve kaynağı farklılık gösterse de et çoğu batı diyetinde birincil bileşen olarak kalmakta ve genellikle sebze garnitürleri ile birlikte sunulmaktadır. Ayrıca et, Türkiye'de gastronomik geleneklerin, kutlamaların veya etkinliklerin bir parçası olarak önemli bir faktör olarak değerlendirilmektedir. Etin biyolojik değerlerinin yanında duyuşal bileşenleri de oldukça fazla miktarda bulunması tüketicilerin tercih sebepleri arasında yer almaktadır Et, tüm gerekli amino asitleri ve birçok vitamin (B vitaminleri, özellikle B12) ve mineral (çinko, demir gibi) sağlayan beslenme açısından mükemmel bir gıdadır. Vücutta kas sentezi ve bakımı için destek sağlar, bu da hem fiziksel işlev hem de metabolik sağlık için önemlidir. Et ayrıca taurin, kreatin, hidroksiprolin, karnosin ve anserin gibi önemli biyolojik aktif bileşenleri de içermektedir. Et tüketimini etkileyen faktörlerin karmaşık etkileşimi göz önüne alındığında, bu çalışma, Türkiye'de et tüketimini makine öğrenimi algoritmaları kullanarak tahmin etmeyi amaçlamaktadır. 1990-2023 yılları arasındaki Gayrisafi Yurt İçi Hasıla (GSYİH), et üretimi, et fiyatları, yem fiyatları, tarımsal GSYİH, nüfus, ithalat ve ihracat verileri, Random Forest, Gradient Boosting, Support Vector Machine, AdaBoost, Neural Network ve Linear Regression modelleri kullanılarak analiz edilmiştir. Sonuçlar, Gradient Boosting ve AdaBoost algoritmalarının en doğru tahminleri sağladığını ve tarımsal GSYİH, et üretimi ve nüfus verilerinin, et tüketimini tahmin etmede önemli olduğunu göstermektedir.

Anahtar Kelimeler: Gastronomik gelenek, et tüketimi, makine öğrenimi, tahmin modelleme, talep tahmini

A Comprehensive Study on Forecasting Meat Consumption Demand in Türkiye Using Machine Learning Algorithms with Data from 1990 to 2023

Abstract

Since ancient times, meat has been a fundamental part of the human diet and continues to be so in many cultures. Despite variations in the amount and source of meat consumed across different countries and cultures, meat remains a primary component in most Western diets, often accompanied by vegetable side dishes. Additionally, meat is considered an important factor in Türkiye as part of gastronomic traditions, celebrations, and events. In addition to its biological content, the significant presence of sensory features in meat is among the reasons for consumer preference. Meat is an excellent food from a nutritional perspective, providing all essential amino acids and many vitamins (B vitamins, particularly B12) and minerals (such as zinc and iron). It supports muscle synthesis and maintenance in the body, which is important for both physical function and metabolic health. Meat also contains important biologically active compounds such as taurine, creatine, hydroxyproline, carnosine and anserine. Given the complex interplay of factors affecting meat consumption, this study aims to estimate and forecast meat consumption in Türkiye using machine learning algorithms. Data from 1990 to 2023, including Gross Domestic Product (GDP), meat production, meat prices, feed prices, agricultural GDP, population, imports, and exports, were analyzed using

¹ Dr. Öğretim Üyesi, Necmettin Erbakan Üniversitesi, Havacılık ve Uzay Bilimleri Fakültesi, Havacılık Elektrik ve Elektronik Bölümü, hakyurek@erbakan.edu.tr, ORCID ID: 0000-0002-0520-9888

² Öğr. Gör. Dr., Necmettin Erbakan Üniversitesi, Meram Meslek Yüksekokulu, Gıda İşleme Bölümü, hkozan@erbakan.edu.tr, ORCID: 0000-0002-2453-1645

Random Forest, Gradient Boosting, Support Vector Machine, AdaBoost, Neural Network, and Linear Regression models. The results indicate that Gradient Boosting and AdaBoost algorithms provided the most accurate predictions, highlighting the importance of agricultural GDP, meat production, and population in forecasting meat consumption.

Keywords: Gastronomic tradition, meat consumption, machine learning, prediction modeling, demand forecasting

1. Giriş

Antik çağlardan beri et, insan diyetinin temel beslenme kaynaklarından biri olmuştur ve hala bu durum devam etmektedir (Oyan vd., 2024). Etin miktarı ve kaynağı ülkeler ve kültürler arasında farklılık gösterse de çoğu Batı ana yemeği bir et içeren yemek içerir ve sebze garnitürleri tamamlayıcıdır. Et, birkaç vitamin ve mineralin yanı sıra tüm gerekli amino asitleri içerir, bu da onu mükemmel bir protein kaynağı yapar (Kushniruk & Rutynskiy, 2022; Tieland vd., 2012).

Et, insan iskelet kasına bileşen olarak eşdeğer olduğu için, kas yapımı ve bakımı için protein sentezini desteklemek için optimal bir kompozisyona sahip amino asitler sağlar. İskelet kas kütlelerinin desteklenmesi ve korunması hem fiziksel işlevin hem de metabolik sağlığın korunması için son derece önemlidir. Bu doğrultuda, yaşlılarda kas gücünün ve kırılabilirliğinin (sarkopeni) yaşa bağlı düşüşlerini önlemek için diyetin önemli bir parçası olan etin alımı, 2-4 yıl boyunca izlenen 1822 yaşlı katılımcının dahil olduğu bir kohort çalışmasında hayvan proteini alımı ile kırılabilirlik insidansı arasında ters bir ilişki gözlemlenmiştir (Sandoval-Insausti vd., 2016). Genç ve fiziksel olarak aktif bireylerde, et proteini alımının vücut kompozisyonu ve kas gücü üzerinde doğrudan faydalı etkileri olduğu belgelenmiştir (Valenzuela vd., 2019). Et, gerekli amino asitlerin kritik bir kaynağı olmasının yanı sıra, önemli biyolojik aktif özelliklere sahip amino asitler, amino asit türevi metabolitler ve peptitler de sağlar. Bu nedenle, taurine, kreatin, hidroksiprolin, karnosin ve anserin gibi maddelerin hepsi esas olarak etten elde edilir ve önemli fizyolojik fonksiyonlar gösterdikleri önerilmiştir (Wu, 2020). Amino asitler, özellikle diğer substratlar mevcut olmadığında, mikrobiyota tarafından olumlu veya olumsuz sağlık etkileri olan metabolitlere fermente edilir; bu fermentasyon özellikle diğer substratlar mevcut olmadığında gerçekleşir. Diyet ve yemeklerin bileşimi bu nedenle bağırsak ortamının önemli belirleyicileridir.

Proteinlerin yanı sıra, et aynı zamanda mineraller ve vitaminler de sağlar. Çinko, hayvansal bazlı yiyeceklerde düşük olan diyetlerde yeterli miktarda tüketilmesi zor olan bir elementtir. Demir, çeşitli yiyeceklerde bol miktarda bulunsa da biyoyararlanımı en yüksek olduğunda kaynağı ettir. Ette, demir kompleks halde ve hem-demir olarak bulunur, bu da non-hem-demirden önemli ölçüde daha yüksek biyoyararlanıma sahiptir. Bu nedenle, ince bağırsakta yaklaşık %23 hem-demir emilirken, bu oran non-hem-demir için yalnızca %2-8'dir (Kalpalathika vd., 1991) ve bu nedenle kırmızı et, en iyi diyet demir kaynağı olarak kalır (Czerwonka & Tokarz, 2017). Vitaminlerle ilgili olarak, et, kompleks B vitaminlerinin önemli bir kaynağıdır. Aslında, et, balık ve diğer hayvansal türevli yiyecekler (örneğin süt ürünleri) doğal olarak B12 vitamini sağlayan tek fermante edilmemiş yiyeceklerdir (Wyness vd., 2011). Bu durum, dengeli bir omnivor diyeti vegan bir diyetle değiştirmenin mineral ve vitamin durumu üzerindeki derin etkilerini düşünme ihtiyacını vurgular.

Danimarka Ulusal Diyet ve Fiziksel Aktivite Anketi 2001-2013'te, et ve et ürünlerinin (tavuk ve balık hariç) ortalama bir Danimarkalı'nın protein (%27), yağ (%21), doymuş yağ asitleri (%20), tekli doymamış yağ asitleri (%26), vitamin A (%40), vitamin D (%16), tiamin (%33), riboflavin (%17), niyasin (%27), vitamin B6 (%21), vitamin B12 (%35), fosfor (%15), demir (%20), çinko (%33) ve selenyum (%25) alımına (toplam alımın % olarak) önemli ölçüde katkıda bulunduğu gösterilmiştir (Pedersen vd., 2015). Etin diyet besin alımına katkısı erkeklerde

kadınlardan daha yüksektir (Biltoft-Jensen vd., 2016; Knudsen vd., 2012). Sadece besleyicilik açısından değil, aynı zamanda lezzeti konusunda da oldukça önemli bir besin kaynağı olan etin, 22’den fazla endüstriyel et ürünü olarak tüketime sunulmaktadır(Tosun & Demirbaş, 2012). Bilindiği üzere gastronomik geleneklerde et ve et ürünleri vazgeçilmez bir parçadır.Örneğin Konya mutfağında ana yemeklerin incelendiği bir çalışmada fırın kebabı, tirit, etli yaprak sarma, ekmek salması, patlıcan orta, kayısıli yahni, erikli yahni, etli düğün pilavı ve etlietmek ve türevleri başta olmak üzere birçok geleneksel yemekte et kullanılmaktadır (Seçim & Akyol, 2022). Ülkemizde ekonomik olarak en değerli beslenme kaynağı olan et ve et ürünlerinin miktarı, artan popülasyon ile birlikte yeterli gelmediği bildirilmiştir (Gürer, 2021). Bu konuda yalnızca tüketim ihtiyacı değil, aynı zamanda hayvancılık ile uğraşan insanlar için de arz çok önemlidir. Nitekim hayvancılık sektörü ayrıca yüz milyonlarca insana gelir ve istihdam kaynağı sağlamaktadır (Herrero vd., 2009; Mehrabi vd., 2020). Hayvancılıktan elde edilen gıda talebinin insan ve doğal sistemler üzerinde yaygın ve çok boyutlu etkisi vardır. Bu etkiler göz önüne alındığında, tarımsal gayrisafı milli hasıla, et üretim miktarı, et fiyatı, yem fiyatı, gayrisafı milli hasıla, nüfus, ithalat ve ihracat gibi talep faktörlerinin birbirleriyle nasıl etkileşime girdiği önem arz eder. Aynı zamanda hayvancılıktan elde edilen gıda talebine nasıl katkıda bulunduğunu anlamak da önemlidir.

Bu gibi çoklu faktörlerin olduğu kompleks problemlerde, ihtiyaç duyulan sonucu elde etmek için birçok farklı yöntem geliştirilmiş ve geliştirilmeye devam etmektedir. Ancak genelde basit olması, çabuk sonuç vermesi, yüksek kaynaklara ihtiyaç duymaması ve ayrıca çok fazla bilgi gerektirmemesi sebebi ile öncelikle lineer bir çözüm denenmektedir. Bu sebeple geleneksel olarak regresyona dayalı tahminlemeler yapılmaktadır. Oysaki geçtiğimiz 20 yıl içerisinde, teknolojik imkanların da artması ile makine öğrenmesi üzerinde doğruluk oranı çok daha yüksek olan ve hata payını minimize eden metotlar geliştirilmiştir. Bu çalışma kapsamında yalnızca Elsevier firması tarafından son on yılda yayınlanan makale sayıları Şekil 1’de verilmiştir. Görüldüğü üzere, bu konuda bilim insanları tarafından özel bir trend oluşmuş ve bu trend hızla artmaktadır.

Şekil 1: 2013-2023 Yılları Arasında Elsevier Tarafından Yayınlanan Makine Öğrenimi Konulu Makale Sayıları



Bu çalışmada olduğu gibi doğrusal olmayan ilişkileri ve karmaşık etkileşimleri çözmek üzere farklı matematiksel modeller geliştirilmiştir. Bu modelleri uygulayan ve hata payını minimize eden bilgisayar sistemleri ile uygulanan makine öğrenimi modelleri, bu gibi karmaşık sorunlara uygun çözüm olarak düşünülebilir. Örneğin bir restorana ya da mağazaya girecek olan müşteri sayısını hesaplamada Boosted Decision Tree Algoritmasını, Decision Forest Regression algoritmasını ve Bayesian Linear Regression algoritmasını kullanarak %85 doğruluğa ulaşımlardır (Tanizaki vd., 2019). Makine öğrenimi ile görüntü işleme yapılarak yemek içeriklerinin de tespit edilebildiği bildirilmiştir (Özel vd., 2023). Ayrıca bir başka çalışmada (Tanizaki vd., 2020), restoranlarda müşterilerin taleplerinin nasıl yönetilmesi gerektiği ve oluşabilecek talebi, müşteri sipariş miktarı ve stok sipariş miktarını tahmin etmek için Random Forest Regression makine öğrenimi kullanmışlar, restoran lokasyon verisi, hava durumu, yapılan etkinlikler gibi dış faktör verileri ile %70'e yakın doğruluk oranını yakalamışlardır. Ancak çalışma bir yıllık veri ile yürütüldüğü için daha yüksek tahminleme doğruluğuna ulaşamamışlardır. Artan gıda talebinin ülke ekonomisi ile ilişkisi üzerine yapılan bir çalışmada (Bozkir & Sezer, 2011), 3 farklı makine öğrenimi denemiştir ve %80'e kadar doğruluk oranı yakalanmıştır. Restoran talebi tahmininde verilerin, makine öğreniminin ve derin öğrenmenin değeri üzerine yapılan bir başka güncel çalışmada (Chae vd., 2024), gelir düzeyi verileri, demografik veriler, tüketici yönelim verileri, enerji ücretleri, hava şartları, sağlık standartları, mevzuat, kanun ve politik veriler ile lokasyon, menü kategorisi, menü ücretleri, daha önceki satışlar, promosyonlar ve reklam gibi oldukça fazla farklı veri kümesinde, makine öğrenimi algoritmalarından Lasso ve Ridge, Random Forest (RF), Gradient Boosting Machine (GBM), Light GBM, eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Deep Auto Regressive (DeepAR), Temporal Fusion Transformer (TFT) ve Neural Hierarchical Interpolation for Time Series (N-HiTS) kullanmışlar ve %90 doğruluğa kadar ulaşabilmişlerdir.

Bu çalışma, beslenme açısından ve hayvancılık sektöründe çalışanlar ile bu konuda yapılması planlanan et politikaları açısından oldukça değerli olan et arzının en büyük indeksi olan tüketim miktarını tahminlemeyi amaçlamaktadır. Bunun için 1990-2023 yılları arasında tarımsal gayrisafi milli hasıla, et üretim miktarı, et fiyatı, yem fiyatı, gayrisafi milli hasıla, nüfus, ithalat ve ihracat verileri kullanılmıştır. Çalışmada yaygın ve başarılı makine öğrenim modelleri olan Random Forest algoritması, Gradient Boosting algoritması, Support Vector Machine algoritması, AdaBoost algoritması, Neural Network algoritması ve geleneksel olarak kullanılan Linear Regression yöntem ve tekniklerinden faydalanılmıştır.

2. Materyal ve Metotlar

2.1. Veri ön hazırlama işlemleri










Bu çalışmada, çalışmanın ana verileri 1990-2023 yılları arasında tarımsal gayrisafi milli hasıla, et üretim miktarı, et fiyatı, yem fiyatı, gayrisafi milli hasıla, nüfus, ithalat ve ihracat verileridir. Bu kapsamda verilerin bir kısmı Türkiye İstatistik Kurumu (<https://data.tuik.gov.tr/>) tarafından sağlanmış, yem ile ilgili veriler ise Türkiye Yem Sanayicileri Birliği (TUYEM), 1996-2000 yılları arası et fiyat bilgileri ise Türkiye Büyük Millet Meclisi tutanaklarından (<https://tbmm.gov.tr/>) elde edilmiştir. Tüm fiyat bilgileri döneminin yıllık ortalama USD kuru ile Amerikan Doları olarak verilmiştir. Tarımsal Gayrisafi Milli Hasıla (TGSMH), Gayrisafi milli hasılda tarıma dayalı verilen oran ile belirlenmektedir. Tarım sektörü, bir ülkenin gıda üretiminin ana kaynağıdır ve bu nedenle tarımsal GSMH, et üretim ve tüketimini doğrudan etkiler. Yüksek tarımsal GSMH, daha fazla kaynak ayrılması ve daha yüksek üretim kapasitesi anlamına gelir, bu da et tüketimini artırabileceği öngörülmüştür. Et tüketiminin en doğrudan etkileyen faktör doğal olarak et üretim miktarı ve fiyatıdır. Üretim miktarındaki artış, pazarda daha fazla et bulunmasını sağlar ve bu da tüketimi artırır. Et fiyatları, tüketici davranışlarını doğrudan etkiler. Yüksek et fiyatları, tüketicilerin et alım gücünü düşürebilir ve bu da tüketimi

azaltabilir. Tam tersine, düşük et fiyatları tüketimi teşvik edebilir. Üretimde önemli bir etkende yem fiyatıdır. Yem fiyatları, hayvancılık maliyetlerini etkileyen önemli bir faktördür. Yüksek yem fiyatları, et üretim maliyetlerini artırarak et fiyatlarını yükseltebilir ve bu da tüketimi olumsuz yönde etkileyebilir. Gayrisafi milli hasıla (GSMH), genel ekonomik büyüme ve refah düzeyi temsiliyetini oluşturur. Tüketici harcamalarını ve dolayısıyla et tüketimini etkileyen önemli bir göstergedir. Yüksek GSMH, daha fazla harcanabilir gelir anlamına gelir ve bu da et tüketimini artırabilir. Et tüketiminde arzın temelini oluşturan faktör nüfustur. Nüfus büyüklüğü ve demografik özellikler, toplam et talebini belirleyen temel faktörlerdir. Daha büyük bir nüfus, genellikle daha yüksek et tüketimi anlamına gelir. İthalat ve ihracat verileri de et tüketimini dolaylı etki gösterebileceği düşünülmüştür. Ülke genelinde yapılan ithalat ve ihracat, iç piyasada et arzını ve fiyatlarını etkileyebilmektedir. İthalat, yerel arzı artırarak fiyatları düşürebilir ve tüketimi artırabilirken, ihracat, yerel arzı azaltarak fiyatları yükseltebilir ve tüketimi kısıtlayabilir.

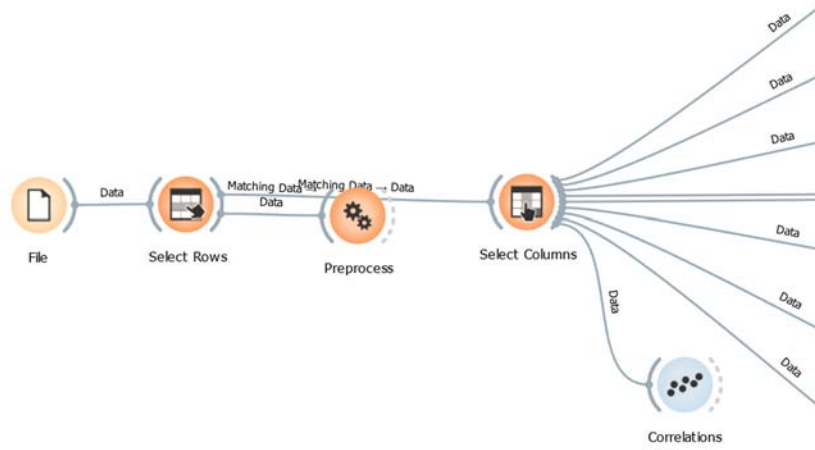
Bu çalışmada, yukarıda belirtilen 8 faktörlerin et tüketimi üzerindeki etkilerini kapsamlı bir şekilde incelenmiştir. Bu 8 faktör, 34 yıllık veriler ile giriş parametrelerini oluşturarak, her bir faktörün et tüketimi üzerindeki potansiyel etkilerini değerlendirmek amacıyla detaylı makine öğrenim analizleri ve istatistiksel analizler yapılmıştır. Girdi ver hedef verilere ilişkin yapılan istatistik analiz sonuçları Tablo 1’de verilmiştir.

Değişkenler arasında istatistiksel bir korelasyon olup olmadığını göstermek için Orange3 yazılımının korelasyon algoritması kullanılarak Spearman ve Pearson korelasyon verileri elde edilmiştir.

Tablo 1: Girdi ve hedef parametrelerine ilişkin tanıttıcı istatistikler

Değişken	N	Dağılım	Ort	SE Mean	Variance	Min.	Maks.	Skewness	Kurtosis
Tarımsal GSYİH	34		54,51	5,43	1002,29	13,16	110,68	0,00	-1,64
Et Üretim Miktarı (bin ton)	34		969,00	93,30	295962,50	408,70	2384,00	1,12	0,45
Et Fiyatları (\$/kg)	34		6,29	0,42	5,90	2,35	12,22	0,41	-0,25
Yem Fiyatları (\$/kg)	34		0,25	0,01	0,01	0,11	0,40	0,29	-0,37
GSYİH (milyar \$)	34		545,30	54,40	100783,40	131,60	1118,60	0,01	-1,63
Nüfus(milyon)	34		71,44	1,44	70,92	56,00	85,00	-0,17	-0,92
İthalat (milyar \$)	34		144,20	17,80	10750,40	21,10	363,70	0,38	-0,96
İhracat (milyar \$)	34		98,90	12,90	5666,00	13,00	255,40	0,47	-0,92
OECD Et Tüketimi (bin ton)	34		1039,70	74,80	190306,30	635,00	2079,00	1,26	0,25

Korelasyon değerini ölçmek için oluşturulan algoritmanın görüntüsü Şekil 2’de gösterilmiştir.



Şekil 2. Orange3 yazılımı tarafından Pearson ve Spearman korelasyon ölçümü için kullanılan algoritma

Korelasyon analizine ilişkin veriler Tablo 2’de verilmiştir. Korelasyon analizi sonuçlarına göre, OECD et tüketimi (bin ton) ile çeşitli ekonomik ve demografik faktörler arasında güçlü ve anlamlı ilişkiler tespit edilmiştir. Pearson ve Spearman korelasyon katsayılarına göre, en yüksek korelasyon et üretim miktarı ile gözlemlenmiştir (Pearson: +0.981, Spearman: +0.973). Bu, et üretimindeki artışın et tüketimini büyük ölçüde artırdığını göstermektedir. İhracatın et tüketimi ile olan ilişkisi de oldukça güçlüdür (Pearson: +0.902, Spearman: +0.939), bu da ihracattaki artışın tüketimi olumlu yönde etkilediğini ifade eder.

Tablo 2. Değişkenler arası Pearson ve Spearman korelasyon analizi

Birinci Değişken	İkinci Değişken	Korelasyon Katsayısı (Pearson)	Korelasyon Katsayısı (Spearman)
Et Üretim Miktarı (bin ton)		+0.981	+0.973
İhracat (milyar \$)		+0.902	+0.939
Nüfus (milyon)		+0.848	+0.932
İthalat (milyar \$)	OECD Et Tüketimi (bin ton)	+0.837	+0.900
GSYİH (milyar \$)		+0.734	+0.862
Tarımsal GSYİH (milyar \$)		+0.733	+0.861
Yem Fiyatları (\$/kg)		+0.360	+0.574
Et Fiyatları (\$/kg)		+0.175	+0.447

Nüfusun et tüketimi ile olan korelasyonu da yüksek seviyededir (Pearson: +0.848, Spearman: +0.932). Bu sonuç, nüfus artışının et tüketimini artırdığını göstermektedir. İthalat da benzer şekilde güçlü bir pozitif korelasyon göstermektedir (Pearson: +0.837, Spearman: +0.900), bu da ithalatın et arzını artırarak tüketimi desteklediğini ifade eder.

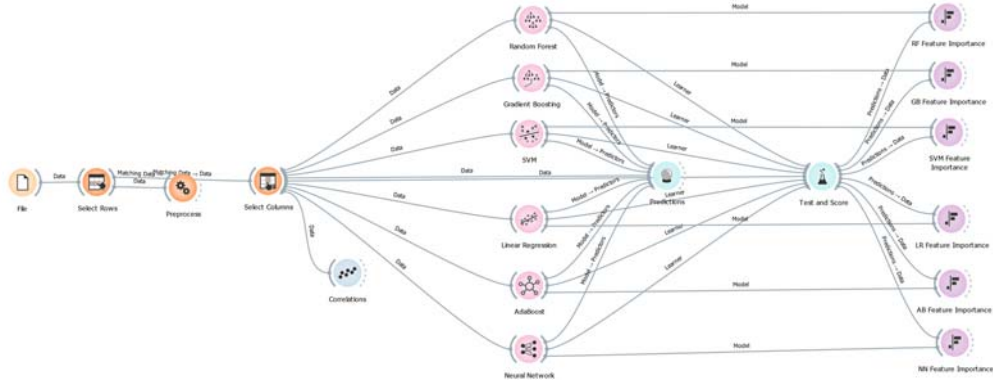
Genel GSYİH ve tarımsal GSYİH ile et tüketimi arasında orta düzeyde pozitif korelasyonlar bulunmuştur (GSYİH Pearson: +0.734, Spearman: +0.862; tarımsal GSYİH Pearson: +0.733, Spearman: +0.861). Bu, ekonomik büyümenin ve tarımsal sektördeki gelişmelerin et tüketimini artırdığını göstermektedir.

Yem fiyatları ile et tüketimi arasında zayıf bir pozitif korelasyon bulunmuştur (Pearson: +0.360, Spearman: +0.574), bu da yem fiyatlarındaki artışın et tüketimini sınırlı düzeyde etkilediğini göstermektedir.

Et fiyatları ile et tüketimi arasındaki korelasyon ise oldukça zayıftır (Pearson: +0.175, Spearman: +0.447), bu da et fiyatlarının tüketim üzerindeki etkisinin sınırlı olduğunu göstermektedir.

2.2. Makine Öğrenim Algoritmaları

Günümüzde makine öğrenimi algoritmaları, karmaşık ve büyük ölçekli veri kümeleri içeren sorunlarda tahmin verisi oluşturmak için en çok tercih edilen yöntemlerdir. Bu çalışmada, Türkiye’deki et tüketimini etkileyen faktörleri dikkate alarak et tüketimini tahmin etmek için makine öğrenimi algoritmaları kullanılmıştır. Bu çalışmada RF (Random Forest), Gradient Boosting, SVM (Support Vector Machine), Linear Regression, AdaBoost ve Neural Network gibi farklı makine öğrenimi algoritmaları Orange yazılımında uygulanmıştır. Her model, çıktıyı farklı algoritmalarla tahmin etmeye çalışmaktadır. Orange yazılımı, GNU Genel Kamu Lisansı (<https://www.gnu.org/licenses/gpl-3.0.html>) altında sağlanan, veri görselleştirme, makine öğrenimi ve veri madenciliği gerçekleştiren açık kaynaklı ve ücretsiz bir bilgisayar programıdır. Bu yazılım, çalışmada kullanılan altı farklı makine öğrenimi modelinin tahmin verilerini tek bir arayüzde hesaplama imkânı sunmuştur. Çalışma için oluşturulan program arayüzü görüntüsü Şekil 3’te sunulmuştur. Bu çalışmada kullanılan makine öğrenimi modellerinin temel tanımları, çalışma prensipleri ve algoritmalarının özetleri ilgili alt bölümlerde kısaca verilmiştir.



Şekil 3. Orange3 yazılımı tarafından makine öğrenim modellerine ilişkin arayüz

Random Forest

Random Forest, karar ağaçlarının bir topluluğudur ve her bir ağaç, veri setinin rastgele bir alt kümesi üzerinde eğitilir. Sonuç olarak, modelin tahmini, tüm ağaçların tahminlerinin çoğunluk oyu veya ortalaması alınarak elde edilir. Bu yöntem, overfitting’i azaltarak genelleme yeteneğini artırır. Her bir karar ağacı, veri setinin bir bootstrapped örneği üzerinde eğitilir ve her bölme noktasında özelliklerin rastgele bir alt kümesi seçilir. Sonuçlar, bagging (bootstrap aggregating) yöntemi ile birleştirilir. Tahmin \hat{y} , B sayıda karar ağacının tahminlerinin ortalaması olarak hesaplanır:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B \hat{y}_b \quad (1)$$

Gradient Boosting

Gradient Boosting, zayıf tahminleyicilerin (genellikle karar ağaçları) bir topluluğunu iteratif olarak eğiten bir tekniktir. Her yeni ağaç, önceki ağaçların hatalarını düzeltmeye çalışır. Model, her iterasyonda hata fonksiyonunun negatif gradyanı doğrultusunda güncellenir. Her yeni ağaç, mevcut modelin hatalarını tahmin etmeye çalışır. Modelin tahmini \hat{y} , n sayıda modelin ağırlıklı toplamı olarak hesaplanır:

Modelin tahmini \hat{y}

$$\hat{y} = \sum_{t=1}^n \gamma_t h_t(x) \quad (2)$$

Burada γ_t , t . zayıf modelin ağırlığı ve $h_t(x)$, t . zayıf modelin tahminidir.

Support Vector Machine (SVM)

SVM, sınıflandırma ve regresyon analizinde kullanılan güçlü bir algoritmadır. Temel amacı, veri noktalarını iki sınıf arasında en geniş marjinle ayıran bir hiper düzlem bulmaktır. SVM, veriyi yüksek boyutlu bir uzaya projeler ve bu uzayda lineer bir ayırma hiper düzlemi bulur. Veri noktalarını bu hiper düzlemden en uzak mesafede tutmaya çalışır.

Hiper düzlem denklemi:

$$f(x) = wx + b \quad (3)$$

Hata fonksiyonu (Hinge loss)

$$\min_{w \in \mathbb{R}^d} \|w\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i f(x_i)) \quad (4)$$

Linear Regression

Lineer regresyon, bağımsız değişkenler ile bağımlı değişken arasındaki lineer ilişkiyi modellemek için kullanılır. Model, bağımsız değişkenlerin bir lineer kombinasyonu olarak bağımlı değişkeni tahmin eder.

Model denklemi:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (5)$$

Burada y bağımlı değişkeni, x bağımsız değişkeni, β_0 sabiti, β_1 ise x 'in eğimi ve ε ise hatayı ifade etmektedir.

Tahmin edici katsayılar $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

AdaBoost

AdaBoost, zayıf tahminleyicilerin (genellikle karar ağaçları) bir topluluğunu iteratif olarak eğitir. Her iterasyonda, hatalı tahmin edilen örneklere daha fazla ağırlık verilir. Her yeni model, önceki modellerin hatalarını azaltmaya çalışır ve hatalı tahmin edilen örneklere daha fazla ağırlık vererek modelin performansını artırır.

Modelin tahmini \hat{y}

$$\hat{y} = \text{sign}\left(\sum_{t=1}^n \alpha_t h_t(x)\right) \quad (7)$$

Burada α_t , t . zayıf modelin ağırlığı ve $h_t(x)$, t . zayıf modelin tahminidir.

Neural Network

Sinir ağları, biyolojik sinir sistemlerinden ilham alınarak geliştirilmiş ve birçok düğümden (nöron) oluşan katmanlar halinde düzenlenmiş bir modeldir. Her bir nöron, belirli bir aktivasyon fonksiyonu kullanarak girdileri işler ve bir çıkış üretir. Sinir ağları, giriş katmanından başlayarak bir dizi gizli katman üzerinden verileri işler ve sonunda bir çıkış katmanı ile sonuç verir. Her bir bağlantı ağırlıklandırılır ve ağırlıklar geri yayılım algoritması ile güncellenir. Model denklemi:

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (8)$$

Burada x_i i . düğüme gelen değer, b ise bias değeridir. Geri yayılım algoritması:

$$w_i^+ = w_i - \eta \frac{\delta E}{\delta w_i} \quad (9)$$

Burada E toplam hata, w_i i . Düğümün ağırlığı ve η ise öğrenme oranıdır.

2.2. Makine Öğrenim Algoritmaları Performans Metrikleri

Yapılan bu çalışmada et tüketiminin tahmin edilmesi amacıyla altı farklı makine öğrenimi algoritması kullanılmıştır ve bu modelleri karşılaştırmak için bazı performans kriterlerinin hesaplanması gerekmektedir. Bu kriterler arasında kök ortalama kare hata (RMSE), ortalama mutlak hata (MAE), ortalama kare hata (MSE) ve korelasyon katsayısı (R) bulunmaktadır. Modeller, değişkenler arasındaki doğrusal bağlantıyı R^2 (determinasyon katsayısı) ile tespit etmek için bu kriterleri kullanır. MAPE, tahmin verilerindeki hatanın yüzdesini hesaplar. RMSE tahmin hatalarının standart sapmasını döndürür ve bu, değişkenliğin bir göstergesidir. Performans kriterlerinin her birinin denklemleri ve denklemlere ilişkin açıklamalar Tablo 3’te verilmiştir.

Tablo 3. Modellere ilişkin performans testlerinde kullanılan denklemler ve açıklamaları

Kısaltma	Denklemler	Tanım	En iyi	En kötü
R^2	$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (\bar{Y} - Y_i)^2}$	Coefficient of determination, RSS: sum of squared residuals TSS: total sum of squares	+1	$-\infty$
MSE	$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i' - Y_i)^2$	Mean square error	0	$+\infty$
RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i' - Y_i)^2}$	Root mean square error	0	$+\infty$
MAE	$MAE = \frac{1}{n} \sum_{i=1}^n Y_i' - Y_i $	Mean absolute error	0	$+\infty$
MPE	$MPE = \frac{100}{n} \sum_{i=1}^n \frac{Y_i - Y_i'}{Y_i}$	Model predictive error	0	$+\infty, -\infty$

MSE, MAE ve RMSE, makine öğrenimi modellerinde ampirik tahmin değerlerini minimize etmek için kullanılan performans ölçüm araçlarıdır. Yüksek bir R^2 değeri ve düşük RMSE, MSE ve MAE değerlerine sahip bir makine öğrenimi modeli, diğer modellerden daha başarılı kabul edilir.

3. Sonuçlar, Tartışma ve Öneriler

Bu çalışma, ekonomik, sosyal ve çevresel veriler kullanılarak yapılan makine öğrenimi modellemesine dayanmaktadır. Et tüketimini etkileyen çeşitli bağımsız değişkenler göz önüne alındığında, klasik tahmin yöntemlerinin yetersiz kalması nedeniyle, daha sağlam ve hata payı düşük bir tahmin modeli geliştirmek kaçınılmaz hale gelmiştir. Son yıllarda, makine öğrenimi algoritmaları, tahmin süreçlerinde karmaşıklık ve yüksek sapma olasılığına rağmen kullanılmaktadır. Bu algoritmalar enerji, sağlık, ekonomi, tarım ve eğitim gibi birçok alanda yaygın olarak kullanılmaktadır. Bu algoritmaların tercih edilmesinin ana nedeni, istatistiksel yaklaşımları göz ardı ederek giriş parametrelerinin verilerinin nasıl değiştiğini öğrenmeleri ve böylece çıkış parametresi için gerekli veriyi sağlamalarıdır. Bu nedenle, çalışmamızda en çok tercih edilen altı farklı makine öğrenimi algoritması kullanılmıştır. Çalışmamızın sonuçlarında, et tüketimini etkileyen çeşitli faktörlerin et tüketimi üzerindeki etkilerini değerlendirmek amacıyla altı farklı makine öğrenimi algoritması kullanılmıştır. Algoritmaların performansları karşılaştırılarak, en doğru tahminleri hangi modelin sağladığı belirlenmiştir. Bu analizler, et tüketimini artırmak veya dengelemek için politika yapımcılar ve sektördeki paydaşlar için önemli bilgiler sunmaktadır.

Algoritma Performans Karşılaştırmaları

Bu çalışmanın ana teması, et tüketim verilerinin etkileyen ekonomik, sosyal ve çevresel faktörlerin veri kümelerini kullanarak et tüketim miktarını tahmin etmektir. Tahmin verilerini hesaplamak için ML algoritmaları kullanıldı. ML algoritmalarının tahmini değerlerini hesaplamak için düşünülen modeller bazı özelliklere sahip olmalıdır. Algoritmalara ait hiper-özellikler Tablo 4'te verilmiştir. Bu özellikler ağaç sayısı, gizli katman değerleri, gizli katmanlardaki nöron sayısı, test verilerindeki gizli düğüm sayısı ve kayıp fonksiyonu dereceleri olarak tanımlanır.

Tablo 4. Her makine öğrenme algoritması için ilk hiper özelliklerin özeti

Algoritmalar	Hiper-özellikler
Random Forest	Ağaç sayısı: 10 Maksimum dikkate alınan öznelik sayısı: Sınırsız Tekrarlanabilir eğitim: Evet Maksimum ağaç derinliği: Sınırsız Maksimum örnek sayısı ile düğüm bölmeyi durdurma: Sınırsız
Gradient Boosting	Yöntem: Gradient Boosting (scikit-learn) Ağaç sayısı: 1000 Öğrenme oranı: 0.1 Tekrarlanabilir eğitim: Hayır Maksimum ağaç derinliği: 10 Eğitim örneklerinin oranı: 1 Maksimum örnek sayısı ile düğüm bölme durdurma: 20
SVM	SVM türü: SVM, $C=1.0$, $\epsilon=0.1$ Çekirdek: Polinom, $(\text{auto } x \cdot y + 1.0)^3$ Sayısal tolerans: 0.0001 İterasyon sınırı: 10000
Linear Regression	Düzeltilme: yok Kesme noktası: Yok
Neural Network	Gizli katmanlar: 9 Aktivasyon: ReLu

AdaBoost	Çözücü: L-BFGS-B Alfa: 0.3 Maksimum iterasyon: 1000 Tekrarlanabilir eğitim: Hayır Temel tahminleyici:Ağaç Tahminci sayısı: 100 Algoritma (sınıflandırma): Samme.r Kayıp (regresyon): Üstel
----------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tablo 4, her makine öğrenmesi algoritması için başlangıç hiper-özelliklerini özetlemektedir. **Random Forest** algoritması, 10 ağaç ve sınırsız maksimum özellik sayısı ile uygulanmış, maksimum ağaç derinliği ve düğüm bölme sınırları belirlenmemiştir. **Gradient Boosting**, scikit-learn kütüphanesi kullanılarak 1000 ağaç ve 0.1 öğrenme oranıyla uygulanmış, maksimum ağaç derinliği 10 ile sınırlandırılmıştır ve bölme düğümlerindeki maksimum örnek sayısı 20 olarak belirlenmiştir. **SVM (Support Vector Machine)** algoritması, polinomial çekirdek ($auto x \cdot y + 1.0$)³ ile $C=1.0$ ve $\varepsilon=0.1$ parametreleri ile kullanılmış, 0.0001 sayısal tolerans ve 10000 yineleme sınırı ile uygulanmıştır. **Linear Regression** modeli, herhangi bir düzenleme olmadan ve kesme noktası olmadan uygulanmıştır. **Neural Network (Sinir Ağı)** modeli, 9 gizli katman, ReLu aktivasyonu, L-BFGS-B çözücü, 0.3 alfa ve 1000 maksimum yineleme ile kullanılmış, tekrar edilebilir eğitim yapılmamıştır. **AdaBoost** algoritması, temel tahminci olarak ağaç kullanılarak 100 tahminci ile uygulanmış, sınıflandırma algoritması olarak Samme.r ve regresyon kaybı olarak üstel fonksiyon kullanılmıştır. Bu hiper-özellikler, her algoritmanın temel yapılandırılmalarını ve performansını belirlemek için kritik öneme sahiptir.

Makine öğrenimi modellerinin değerlendirilmesinde kullanılan iki yaygın örnekleme türü, "10-kat Çapraz Doğrulama" ve "Shuffle Split" yöntemleridir. 10-kat Çapraz Doğrulama, veri setini on eşit parçaya böler ve her bir parçayı test seti olarak kullanırken kalan dokuz parçayı eğitim seti olarak kullanır. Bu işlem on kez tekrarlanır ve her bir parçanın test seti olarak kullanılması sağlanır. Bu yöntem, modelin farklı veri alt kümeleri üzerindeki performansını değerlendirerek genel performansın daha sağlam ve güvenilir bir şekilde ölçülmesini sağlar (Kohavi, 1995). Yapmış olduğumuz bu çalışmada elde edilen sonuçlar Tablo 5'te verilmiştir.

Tablo 5. 10-kat Çapraz Doğrulama yöntemi ile yapılan örneklemlerde, tahminlemelere ilişkin, 6 farklı makine öğrenimine ait hatalar ve regresyon sonuçları

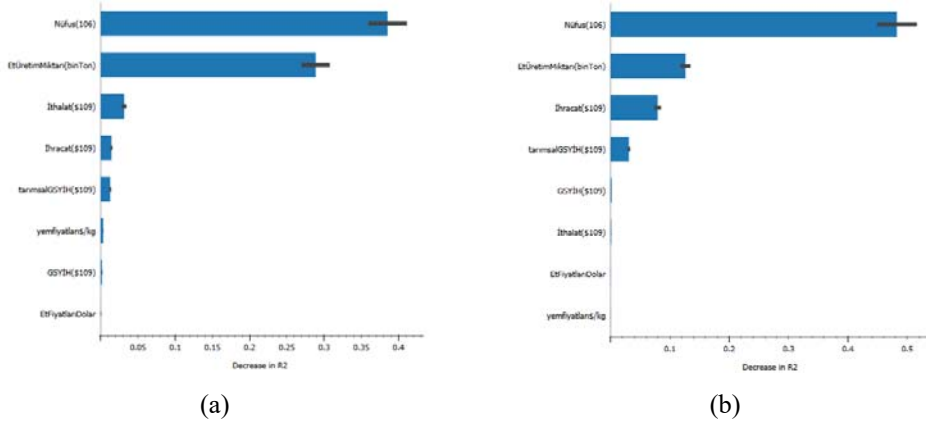
Örnekleme türü: 10-kat Çapraz Doğrulama							
Model	Train	Test	MSE	RMSE	MAE	MAPE	R ²
Random Forest	0,087	0,010	7706,491	87,787	58,656	0,050	0,958
Gradient Boosting	2,080	0,008	3258,624	57,084	41,509	0,041	0,982
SVM	0,027	0,012	168235,213	410,165	249,753	0,175	0,089
Linear Regression	0,015	0,004	10045,438	100,227	78,232	0,081	0,946
AdaBoost	0,644	0,043	4472,647	66,878	44,529	0,042	0,976
Neural Network	0,697	0,010	6891,523	83,015	58,889	0,059	0,963

Öte yandan, Shuffle Split yöntemi, veri setini rastgele bir şekilde bölerek belirli bir yüzdeyi (genellikle %75) eğitim seti, geri kalanını ise test seti olarak ayırır. Bu işlem belirli sayıda kez (bu çalışmada 20 kez) tekrarlanır ve her bir tekrarda farklı rastgele bölünmeler kullanılır. Shuffle Split, özellikle verinin karışık ve düzensiz olduğu durumlarda modelin genel performansını ölçmek için kullanışlıdır (Pedregosa vd., 2011). Bu çalışmada verilerin örnekleme türü olarak Shuffle Split yöntemi seçilerek elde edilen sonuçlar Tablo 6'da verilmiştir.

Tablo 6. Shuffle split (20 rasgele örnek, 75% veri) yöntemi ile yapılan örneklemlerde tahminlemelere ilişkin, 6 farklı makine öğrenimine ait hatalar ve regresyon sonuçları

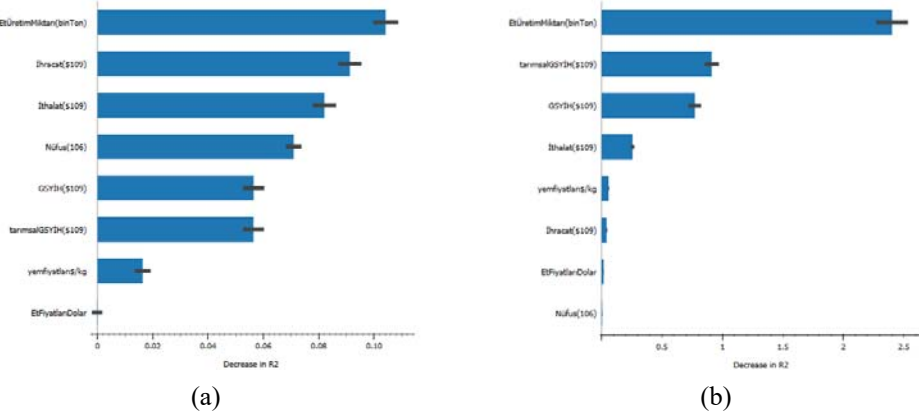
Örnekleme türü: Shuffle split, 20 rasgele örnek, 75% veri							
Model	Train	Test	MSE	RMSE	MAE	MAPE	R ²
Random Forest	0,160	0,014	10166,566	100,829	64,228	0,055	0,945
Gradient Boosting	3,378	0,017	3845,917	62,015	43,840	0,041	0,979
SVM	0,050	0,024	178252,711	422,200	252,219	0,176	0,043
Linear Regression	0,027	0,014	13013,380	114,076	90,237	0,096	0,930
AdaBoost	1,129	0,083	5441,378	73,766	49,567	0,045	0,971
Neural Network	1,362	0,017	12137,274	110,169	72,718	0,071	0,935

Bu çalışmada, çeşitli makine öğrenimi modelleri kullanılarak elde edilen sonuçlar incelenmiştir. En iyi performansı gösteren model Gradient Boosting olmuştur. Elde edilen tüm sonuçlar, et tüketim tahminlemesinde makine öğrenimi yöntemlerinde başarılı sonuçların elde edildiğini göstermiş olup, tahminlemede kullanılan tarımsal gayrisafi milli hasıla, et üretim miktarı, et fiyatı, yem fiyatı, gayrisafi milli hasıla, nüfus, ithalat ve ihracat verileri, et tüketim tahmininde ne kadar önemli olduğunu incelemek de çok önemlidir. Bu kapsamda, Random Forest algoritması ile Gradient Boost algoritması üzerine önemli parametreler Şekil 4'te, SVM algoritması ve Lineer Regresyon Şekil 5'te ve AdaBoost algoritması ile Neural Network algoritmasına ait önemli parametreler ise Şekil 6'da verilmiştir.



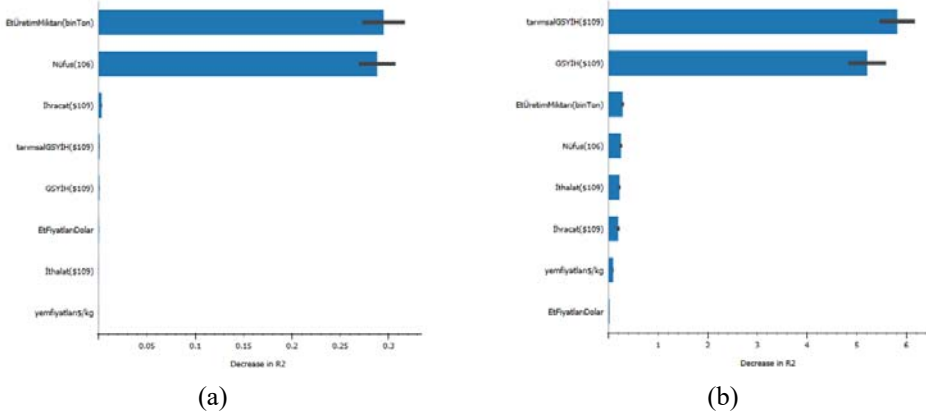
Şekil 4. Algoritmalarda hedeflenen et tüketimi üzerine özelliklerin önem dereceleri
(a) Random Forest algoritması, (b) Gradient Boost algoritması

Random Forest algoritmasında, nüfus ve et üretim miktarı verileri en önemli özellikler olarak öne çıkmıştır. İthalat, ihracat, tarımsal gayrisafi milli hasıla değerleri de önemli etkindir, ancak nüfus ve et üretim miktarının ağırlığı daha yüksektir. Gradient Boost algoritmasında ise benzer şekilde nüfus ve et üretim miktarı verileri en önemli faktörler olarak belirlemiştir. Aynı zamanda ihracat ve tarımsal gayrisafi milli hasıla verileri, daha belirgin düzeyde etki etmişlerdir. Bu durum, Random Forest algoritmasına kıyasla daha az ağaç derinliği ve öğrenme oranı kullanarak daha hassas bir tahmin yeteneği sergilemiştir.



Şekil 5. Algoritmalarda hedeflenen et tüketimi üzerine özelliklerin önem dereceleri (a) SVM algoritması, (b) Linear Regresyon

SVM algoritması, et tüketim tahmininde et fiyatı hariç diğer bütün faktörler önemli özellikler olarak belirlenmiştir. Ancak, genel performansı diğer algoritmalara göre daha düşük kalmıştır. Linear regresyon modelinde beklenildiği üzere et üretimi en önemli faktör olurken, diğer parametreler de modelde düşük ağırlığa sahip kalmıştır. Bu model, doğrusal ilişkileri yakalamada etkili olmasına rağmen, karmaşık ilişkileri ve değişken etkileşimlerini modelleyememiştir.



Şekil 6. Algoritmalarda hedeflenen et tüketimi üzerine özelliklerin önem dereceleri (a) AdaBoost algoritması, (b) Neural Network algoritması

AdaBoost algoritmasında ise et üretim miktarı ve nüfus verileri, en önemli faktörler olarak belirlenmiştir. Neural Network algoritmasında ise tarımsal gayrisafi milli hasıla ile gayrisafi milli hasıla verileri önemli faktörler olarak tanımlanmıştır. Diğer bütün parametreler, modelin tahminlemesinde birer faktör olarak kullanılmış ancak düşük ağırlığa sahip olmuşlardır. Karmaşık veri ilişkilerini modelleme kabiliyeti sayesinde, et tüketim tahminlerinde etkili olmuştur.

Bu çalışmada, çeşitli makine öğrenmesi algoritmalarının—Random Forest, Gradient Boosting, SVM, Linear Regression, AdaBoost ve Neural Network—performansları iki farklı doğrulama yöntemi kullanılarak değerlendirildi: 10-kat Çapraz Doğrulama ve Shuffle Split (%75 veri ile 20 rastgele örnek). Ana hedef, veri setimiz için en doğru tahminleri ve en düşük hata oranlarını sağlayan algoritmaları belirlemektir.

- **Random Forest** algoritması, 10-kat çapraz doğrulamada R^2 değeri açısından oldukça yüksek bir performans sergilemiştir ($R^2 = 0.958$). Shuffle Split yönteminde de benzer bir performans sergileyerek ($R^2 = 0.945$) tutarlılığını korumuştur. Ancak, MSE ve RMSE değerleri diğer modellere göre daha yüksek bulunmuştur. (Fisher vd., 2024), Random Forest algoritmasının genel olarak yüksek doğruluk ve düşük overfitting oranı ile karakterize edildiğini belirtmiştir. Benzer şekilde (Breda vd., 2024) ve (Yang vd., 2024) gibi araştırmacılar, Random Forest'in farklı veri setlerinde yüksek performans sergilediğini göstermiştir.
- **Gradient Boosting** algoritması, her iki doğrulama yönteminde de en düşük MSE ve RMSE değerlerine sahip olmuştur. Bu durum, modelin yüksek doğruluk ve düşük hata oranı ile çalıştığını göstermektedir. Özellikle Shuffle Split yönteminde $R^2 = 0.979$ değeri ile ve 10-kat çapraz doğrulamada $R^2 = 0.982$ değerleri ile en yüksek performansı göstermiştir. (Friedman, 2001), Gradient Boosting'in yüksek doğruluk ve esneklik sunduğunu belirtmiştir. Ayrıca, (Chen & Guestrin, 2016) tarafından geliştirilen XGBoost algoritması, Gradient Boosting'in pratik uygulamalarda nasıl üstün performans gösterdiğini detaylandırmaktadır.
- **SVM** algoritması, her iki doğrulama yönteminde de en düşük R^2 değerlerine sahip olmuştur ($R^2 = 0.089$ ve 0.043). Ayrıca, MSE ve RMSE değerleri de oldukça yüksektir, bu da modelin tahmin performansının düşük olduğunu göstermektedir. (Cortes & Vapnik, 1995), SVM'nin teorik temelini oluşturmuş ve belirli veri setlerinde yüksek doğruluk sağladığını göstermiştir. Ancak, (Bennett & Campbell, 2000) çalışmalarında, SVM'nin karmaşık ve yüksek boyutlu veri setlerinde performansının düşebileceğini belirtmişlerdir. (Huang vd., 2023) uzun kısa süreli hafızalı sinir ağına dayalı eğitim kararlılığı tahmini: evrişimli sinir ağları, SVM ve RF modelleri ile karşılaştırma yapmış, ve SVM'nin düşük çıktığını raporlamış olup, (Tanveer vd., 2023) de yapmış olduğu çalışmada SVM'nin diğer makine öğrenimlerinden düşük çıktığını gözlemlemiştir.
- **Linear Regression** modeli, orta düzeyde performans sergilemiş, ancak Gradient Boosting ve AdaBoost modellerinin gerisinde kalmıştır. 10-kat çapraz doğrulamada $R^2 = 0.946$ iken, Shuffle Split yönteminde $R^2 = 0.930$ bulunmuştur. Seber ve Lee (2012) tarafından yapılan çalışmalar, doğrusal regresyon modellerinin genellikle basit ve hızlı olmasına rağmen, karmaşık veri ilişkilerini modellemekte yetersiz kalabileceğini vurgulamaktadır. Ayrıca, (Draper & Smith, 1998) doğrusal regresyonun sınırlamalarını ve potansiyel hata kaynaklarını detaylandırmaktadır. (Fitri, 2023) yapmış oldukları çalışmada en düşük doğruluğu diğer makine öğrenim yöntemlerine göre lineer regresyonda bulmuş, (Nadkarni vd., 2023) ise RF ve Gradient Boosting algoritmalarını, tahminleme üzerine yaptığı bir çalışmada değerlendirirken, lineer regresyonun bu tip tahminleme çalışmalarında yetersiz kaldığını bildirmişlerdir. (Munir, 2023) Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbor makine öğrenim metodlarını lineer regresyonla kıyaslamış ve en düşük doğruluk değerini diğerlerine göre oldukça düşük sonuç verdiği şekliyle lineer regresyonda gözlemlemiştir.
- **AdaBoost** algoritması, 10-kat çapraz doğrulamada oldukça iyi bir performans sergilemiştir ($R^2 = 0.976$). Shuffle Split yönteminde de benzer şekilde yüksek bir

performans göstermiştir ($R^2 = 0.971$). (Freund & Schapire, 1997)tarafından geliştirilen AdaBoost, zayıf öğrencilerin birleşimi ile güçlü tahminler yapabilme kabiliyeti ile bilinir. Ayrıca, (Sevinç, 2022), (Wang & Sun, 2021) ve (Ding vd., 2022) yapmış oldukları farklı çalışmalarda, AdaBoost’un farklı veri setlerinde nasıl başarılı olduğunu göstermektedir.

- **Neural Network** modeli, her iki doğrulama yönteminde de iyi performans sergilemiştir. 10-kat çapraz doğrulamada $R^2 = 0.963$ iken, Shuffle Split yönteminde $R^2 = 0.935$ bulunmuştur. (Braga-Neto, 2020) ve (Fieguth, 2022) tarafından yapılan çalışmalar, sinir ağlarının karmaşık veri ilişkilerini modelleme kabiliyetini ve yüksek doğrulukta tahminler yapma potansiyelini vurgulamaktadır.

Sonuç olarak, **Gradient Boosting** ve **AdaBoost** algoritmaları, düşük hata oranları ve yüksek doğruluk ile en iyi performansı sergileyen modeller olmuştur. Özellikle Gradient Boosting, her iki doğrulama yönteminde de en düşük MSE ve RMSE değerleri ile dikkat çekmiştir. Random Forest ve Neural Network algoritmaları da yüksek performans göstermiştir. **SVM** ise en düşük performansı gösteren model olarak belirlenmiştir. Bu çalışmanın sonuçları, Gradient Boosting ve AdaBoost algoritmalarının yüksek doğruluk ve düşük hata oranları ile veri tahminlerinde etkili olduğunu göstermektedir. Bu makine öğrenim modellerinde, parametrelerin etkileri de önemlidir. Bu çalışmanın bulguları, tarımsal gayrisafı milli hasıla, et üretim miktarı ve nüfus parametrelerinin, et tüketim tahminlerinde kritik rol oynadığını göstermektedir. Özellikle Gradient Boost ve Random Forest algoritmaları, bu özellikleri etkili bir şekilde kullanarak yüksek doğrulukta tahminler sağlamıştır. Diğer yandan, SVM ve Lineer Regresyon gibi modeller, daha düşük performans sergilemiş ve karmaşık ilişkileri modelleme yeteneklerinin sınırlı olduğunu göstermiştir. AdaBoost ve Neural Network modelleri ise, özellikle et üretim miktarı ve nüfus gibi değişkenlerin önemini vurgulayarak, tahminlerde başarılı olmuştur.

Bu bulgular, mevcut literatürdeki diğer çalışmalarla uyum göstermektedir. Örneğin, (Breda vd., 2024), (Yang vd., 2024) veya (Fisher vd., 2024) in yaptıkları çalışmalarda, ensemble yöntemlerinin (Random Forest ve Gradient Boosting gibi) yüksek doğrulukta tahminler sağladığını ve veri setlerindeki önemli özellikleri etkili bir şekilde belirleyebildiğini göstermektedir.

Bu konularda ileriki çalışmalar için bazı öneriler sunulmuştur. Bulgularımızın sağlamlığını ve uygulanabilirliğini daha da artırmak üzere grid search veya random search yöntemleri kullanılarak kapsamlı hiperparametre ayarlamaları model performansını potansiyel olarak artırabilir. Verideki daha alakalı desenleri ve ilişkileri yakalamak için ek özellik mühendisliği tekniklerinin araştırılması, model doğruluğunu artırabilir. Stacking, Blending veya Bagging gibi ileri seviye ensemble teknikleri araştırılarak, birden fazla modelin güçlü yönlerini birleştirip daha iyi tahmin performansı elde edilebilir. Ek harici veri setlerini bütünleştirerek, daha fazla bağlam veya ek özellikler sağlayarak modelin doğruluğu artırılabilir. Bu alanlarda yapılacak ileriki çalışmalarla, modellerin daha da iyileştirip tahmin doğruluğunun artacağı ve verideki temel desenlerin daha derinlemesine anlaşılacağı değerlendirilmiştir.

Etik Kurul Onayı

Bu çalışma için Etik kurul kararı gerekmemektedir.

Yazar Katkıları

Araştırma Tasarımı (CRediT 1) Yazar 1 (%50) – Yazar 2 (%50)

Veri Toplama (CRediT 2) Yazar 1 (%50) – Yazar 2 (%50)

Araştırma - Veri Analizi - Doğrulama (CRediT 3-4-6-11) Yazar 1 (%50) – Yazar 2 (%50)

Makalenin Yazımını (CRediT 12-13) Yazar 1 (%50) – Yazar 2 (%50)

Metnin Tashihi ve Geliştirilmesi (CRediT 14) Yazar 1 (%50) – Yazar 2 (%50)

Finansman

Çalışma herhangi bir finansman desteği almamıştır.

Çıkar Çatışması

Yazarlar ya da üçüncü taraflar açısından çalışmadan kaynaklı çıkar çatışması bulunmamaktadır.

Sürdürülebilir Kalkınma Amaçları (SDG)

Sürdürülebilir Kalkınma Amaçları: Desteklemiyor

REFERANSLAR

- Bennett, K. P., & Campbell, C. (2000). Support vector machines. *ACM SIGKDD explorations newsletter*, 2(2), 1-13. <https://doi.org/10.1145/380995.380999>
- Biltoft-Jensen, A. P., Kørup, K., Christensen, T., Eriksen, K., Ygil, K. H., & Fagt, S. (2016). *Køds rolle i kosten*. DTU Fødevareinstituttet.
- Bozkir, A. S., & Sezer, E. A. (2011). Predicting food demand in food courts by decision tree approaches. *Procedia Computer Science*, 3, 759-763. <https://doi.org/10.1016/j.procs.2010.12.125>
- Braga-Neto, U. (2020). *Fundamentals of pattern recognition and machine learning*. Springer. <https://doi.org/10.1007/978-3-030-27656-0>
- Breda, L. S., de Melo Nascimento, J. E., Alves, V., de Toledo, V. d. A. A., de Lima, V. A., & Felsner, M. L. (2024). Green and fast prediction of crude protein contents in bee pollen based on digital images combined with Random Forest algorithm. *Food Research International*, 179, 113958. <https://doi.org/10.1016/j.foodres.2024.113958>
- Chae, B., Sheu, C., & Park, E. O. (2024). The value of data, machine learning, and deep learning in restaurant demand forecasting: Insights and lessons learned from a large restaurant chain. *Decision Support Systems*, 114291. <https://doi.org/10.1016/j.dss.2024.114291>
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297. <https://doi.org/10.1007/BF00994018>

Akyürek, H.A., Kozan, H.İ. (2024). “Makine öğrenmesi algoritmalarını kullanarak 1990-2023 yılları arası veriler ile Türkiye’de et tüketim talebinin tahmini üzerine kapsamlı bir çalışma”, *Necmettin Erbakan Üniversitesi NEUGastro Dergisi*, 3(2), 1 – 19.
Doi:

Başvuru : 30.07.2024 Kabul : 11.09.2024 Araştırma Makalesi/Research Article

- Czerwonka, M., & Tokarz, A. (2017). Iron in red meat–friend or foe. *Meat Science*, 123, 157-165. <https://doi.org/10.1016/j.meatsci.2016.09.012>
- Ding, Y., Zhu, H., Chen, R., & Li, R. (2022). An efficient AdaBoost algorithm with the multiple thresholds classification. *Applied sciences*, 12(12), 5872. <https://doi.org/10.3390/app12125872>
- Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). John Wiley & Sons.
- Fieguth, P. (2022). *An introduction to pattern recognition and machine learning*. Springer. <https://doi.org/10.1007/978-3-030-95995-1>
- Fisher, J., Allen, S., Yetman, G., & Pistolesi, L. (2024). Assessing the influence of landscape conservation and protected areas on social wellbeing using random forest machine learning. *Scientific Reports*, 14(1), 11357. <https://doi.org/10.1038/s41598-024-61924-4>
- Fitri, E. (2023). Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah. *Journal of Applied Computer Science and Technology*, 4(1), 58-64. <https://doi.org/10.52158/jacost.v4i1.491>
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Gürer, B. (2021). Türkiye’de nüfusun yeterli ve dengeli beslenmesi açısından hayvansal gıda arz ve talebinin değerlendirilmesi. *Gıda*, 46(6), 1450-1466. <https://doi.org/10.15237/gida.GD21083>
- Herrero, M., Thornton, P. K., Gerber, P., & Reid, R. S. (2009). Livestock, livelihoods and the environment: understanding the trade-offs. *Current Opinion in Environmental Sustainability*, 1(2), 111-120. <https://doi.org/10.1016/j.cosust.2009.10.003>
- Huang, F., Xiong, H., Chen, S., Lv, Z., Huang, J., Chang, Z., & Catani, F. (2023). Slope stability prediction based on a long short-term memory neural network: Comparisons with convolutional neural networks, support vector machines and random forest models. *International Journal of Coal Science & Technology*, 10(1), 18. <https://doi.org/10.1007/s40789-023-00579-4>
- Kalpalathika, P. M., Clark, E. M., & Mahoney, A. W. (1991). Heme iron content in selected ready-to-serve beef products. *Journal of Agricultural and Food Chemistry*, 39(6), 1091-1093. <https://doi.org/10.1021/jf00006a017>
- Knudsen, V., Fagt, S., Trolle, E., Matthiessen, J., Groth, M., Biloft-Jensen, A., Sørensen, M., & Pedersen, A. (2012). Evaluation of dietary intake in Danish adults by means of an index based on food-based dietary guidelines. *Food & nutrition research*, 56(1), 17129. <https://doi.org/10.3402/fnr.v56i0.17129>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*,

- Kushniruk, H., & Rutynskyi, M. (2022). Development of the infrastructure of Turkish restaurants in the tourist center of Eastern Europe: the case of Kyiv. *GastroMedia Journal*, 1(1), 1-18. <https://doi.org/10.54497/Gastromedia.2022.1>
- Mehrabi, Z., Gill, M., Wijk, M. v., Herrero, M., & Ramankutty, N. (2020). Livestock policy for sustainable development. *Nature Food*, 1(3), 160-165. <https://doi.org/10.1038/s43016-020-0042-9>
- Munir, M. D. (2023). Prediction of Heteroscedastic Data Using Linear Regression and Various Machine Learning Models. *Int. J. Sci. Res. in Mathematical and Statistical Sciences Vol*, 10(1).
- Nadkarni, S. B., Vijay, G., & Kamath, R. C. (2023). Comparative Study of random forest and gradient boosting algorithms to predict airfoil self-noise. *Engineering Proceedings*, 59(1), 24. <https://doi.org/10.3390/engproc2023059024>
- Oyan, O., Şenyüz, H., & Arköse, C. (2024). Comparison of carcass weight and carcass characteristics in some cattle breeds. *Research and Practice in Veterinary and Animal Science (REPVAS)*, 1(1), 1-8. <https://doi.org/10.69990/repvas.2024.1.1.1>
- Özel, M. A., Gül, M. Y., & Güneş, E. (2023). YOLO Algoritması ile Gastronomide Yemek İçeriklerini Tespit Etmede Bir Makine Öğrenimi Yaklaşımı. *Necmettin Erbakan Üniversitesi GastroMedia Dergisi (NEU GastroMedia)*, 2(1), 59. <https://doi.org/10.54497/Gastromedia.2023.7>
- Pedersen, A. N., Christensen, T., Matthiessen, J., Knudsen, V. K., Sørensen, M. R., Biltoft-Jensen, A. P., Hinsch, H.-J., Ygil, K. H., Kørup, K., & Saxholt, E. (2015). *Danskernes kostvaner 2011-2013*. DTU Fødevareinstituttet.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12, 2825-2830. <https://doi.org/https://dl.acm.org/doi/10.5555/1953048.2078195>
- Sandoval-Insausti, H., Pérez-Tasigchana, R. F., López-García, E., García-Esquinas, E., Rodríguez-Artalejo, F., & Guallar-Castillón, P. (2016). Macronutrients intake and incident frailty in older adults: a prospective cohort study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 71(10), 1329-1334. <https://doi.org/10.1093/gerona/glw033>
- Seçim, Y., & Akyol, N. (2022). *Konya Mutfak Kültürü* (A. Cihan, Ed.). Akademik Kitaplar.
- Sevinç, E. (2022). An empowered AdaBoost algorithm implementation: A COVID-19 dataset study. *Computers & Industrial Engineering*, 165, 107912. <https://doi.org/10.1016/j.cie.2021.107912>
- Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. *Procedia CIRP*, 79, 679-683. <https://doi.org/10.1016/j.procir.2019.02.042>
- Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurants store management based on demand forecasting. *Procedia CIRP*, 88, 580-583. <https://doi.org/10.1016/j.procir.2020.05.101>

Akyürek, H.A., Kozan, H.İ. (2024). “Makine öğrenmesi algoritmalarını kullanarak 1990-2023 yılları arası veriler ile Türkiye’de et tüketim talebinin tahmini üzerine kapsamlı bir çalışma”, *Necmettin Erbakan Üniversitesi NEUGastro Dergisi*, 3(2), 1 – 19.
Doi:

Başvuru : 30.07.2024

Kabul

: 11.09.2024

Araştırma Makalesi/Research Article

- Tanveer, H., Adam, M. A., Khan, M. A., Ali, M. A., & Shakoor, A. (2023). Analyzing the Performance and Efficiency of Machine Learning Algorithms, such as Deep Learning, Decision Trees, or Support Vector Machines, on Various Datasets and Applications. *The Asian Bulletin of Big Data Management*, 3(2), 126-136. <https://doi.org/10.62019/abbdm.v3i2.83>
- Tieland, M., Borgonjen-Van den Berg, K. J., van Loon, L. J., & de Groot, L. C. (2012). Dietary protein intake in community-dwelling, frail, and institutionalized elderly people: scope for improvement. *European journal of nutrition*, 51, 173-179. <https://doi.org/10.1007/s00394-011-0203-6>
- Tosun, D., & Demirbaş, N. (2012). Türkiye’de kırmızı et ve et ürünleri sanayiinde gıda güvenliği sorunları ve öneriler. *Uludağ Üniversitesi Ziraat Fakültesi Dergisi*, 26(1), 93-102.
- Valenzuela, P. L., Mata, F., Morales, J. S., Castillo-García, A., & Lucia, A. (2019). Does beef protein supplementation improve body composition and exercise performance? A systematic review and meta-analysis of randomized controlled trials. *Nutrients*, 11(6), 1429. <https://doi.org/10.3390/nu11061429>
- Wang, W., & Sun, D. (2021). The improved AdaBoost algorithms for imbalanced data classification. *Information Sciences*, 563, 358-374. <https://doi.org/10.1016/j.ins.2021.03.042>
- Wu, G. (2020). Important roles of dietary taurine, creatine, carnosine, anserine and 4-hydroxyproline in human nutrition and health. *Amino acids*, 52(3), 329-360. <https://doi.org/10.1007/s00726-020-02823-6>
- Wyness, L., Weichselbaum, E., O’connor, A., Williams, E., Benelam, B., Riley, H., & Stanner, S. (2011). Red meat in the diet: an update. *Nutrition Bulletin*, 36(1), 34-77. <https://doi.org/10.1111/j.1467-3010.2010.01871.x>
- Yang, X., Hua, Z., Li, L., Huo, X., & Zhao, Z. (2024). Multi-source information fusion-driven corn yield prediction using the Random Forest from the perspective of Agricultural and Forestry Economic Management. *Scientific Reports*, 14(1), 4052. <https://doi.org/10.1038/s41598-024-54354-9>