

Acil serviste kullanılan tıbbi sarf malzemelerinin makine öğrenmesi yöntemleriyle tüketim tahmini: Bir özel hastane örneği¹

Forecasting the consumption of medical consumables used in the emergency department using machine learning methods: A private hospital case

Ercan Çulha¹ 

Abdullah Mısırlıoğlu² 

Refika Sultan Doğan³ 

Öz

Amaç: Bu çalışma, acil serviste kullanılan yüksek tüketimli tıbbi sarf malzemelerinin aylık tüketim miktarlarının makine öğrenmesi yöntemleriyle tahmin edilebilirliğini incelemeyi amaçlamaktadır. Yöntem: Ankara'da özel bir hastanenin 2022-2025 dönemine ait 10 malzemeye ilişkin 480 aylık gözlem kullanılarak altı makine öğrenmesi algoritması, hiperparametre optimizasyonu ile kronolojik ve ay düzeyinde gruplanmış çapraz doğrulama kapsamında karşılaştırılmıştır. En uygun model, yalnızca çapraz doğrulama sonuçlarına dayanan önceden belirlenmiş seçim kuralıyla seçilmiş; final test seti yalnızca bağımsız doğrulamada kullanılmıştır. Hasta başvuru sayısı veri sızıntısını önlemek amacıyla gecikmeli olarak modele alınmıştır. Bulgular: Random Forest en uygun model olarak belirlenmiş (çapraz doğrulama $R^2=0,792$; $SS=0,008$) ve bağımsız test R^2 değeri 0,887 bulunmuştur. Değişken çıkarma analizi, hasta başvuru sayısının Random Forest modeline ölçülebilir bir artımlı katkı sağlamadığını; katkının ağırlıklı olarak doğrusal modellerde gözlemlendiğini göstermiştir. Sonuç: Hasta başvuru yoğunluğu ile toplu tüketim arasında potansiyel bir ilişki bulunmakla birlikte, bu ilişkinin ürün bazındaki ve artımlı katkı analizindeki karşılığı sınırlıdır. Model, toplu bütçeleme ve insan denetimli karar desteğinde kullanılabilir; ancak ürün bazlı otomatik sipariş kararları için tek başına yeterli değildir.

Anahtar Kelimeler: Tıbbi Sarf Malzeme; Makine Öğrenmesi, Tüketim Tahmini

JEL Kodları: I11, M11, C53

Abstract

Objective: This study examines whether monthly consumption of high-volume medical consumables in an emergency department can be forecast using machine learning. Method: Six algorithms were compared using 480 monthly observations for 10 items from a private hospital in Ankara (2022-2025). Hyperparameters were tuned through chronological, month-grouped cross-validation. The best-performing model was selected solely from cross-validation results, while the held-out test set was used only for independent validation. Emergency-department visit volume was included in lagged form. Findings: Random Forest was selected as the best-performing model (cross-validation $R^2=0.792$; $SD=0.008$) and achieved an independent test R^2 of 0.887. Ablation analysis showed no measurable incremental contribution of visit volume to Random Forest, although positive contributions were observed in linear models. Conclusion: The relationship between visit volume and consumption was limited at the product level. The model may support aggregate budgeting and human-supervised decision-making, but it is not sufficient as a standalone tool for automated product-level ordering.

Keywords: Medical Consumables; Machine Learning, Consumption Forecasting

JEL Codes: I11, M11, C53

¹ Bu çalışma, Ercan Çulha tarafından Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü Sağlık Kuruluşları Yöneticiliği Anabilim Dalı'nda, Dr. Abdullah Mısırlıoğlu danışmanlığında hazırlanmakta olan "Özel Hastanelerde Yapay Zeka Yöntemiyle ve Makine Öğrenimi Tabanlı Talep Tahmini: Ankara İli Örneği" başlıklı doktora tezinden üretilmiştir.

¹ Sivas Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sağlık Kuruluşları Yöneticiliği Doktora Programı, Sivas, Türkiye, culhae.55@gmail.com
ORCID: 0000-0001-6932-4917

² Dr. Öğr. Üyesi, Sivas Cumhuriyet Üniversitesi, Sağlık Bilimleri Fakültesi, Sağlık Yönetimi Bölümü, Sivas, Türkiye
E-posta: misirlioglu@cumhuriyet.edu.tr
ORCID: 0000-0001-8520-7403

³ Dr., Abdullah Gül Üniversitesi, Yaşam ve Doğa Bilimleri Fakültesi, Biyomühendislik Bölümü, Kayseri, Türkiye
E-posta: refikasultan.dogan@agu.edu.tr
ORCID: 0000-0001-8416-1765

Sorumlu Yazar/Corresponding Author:

Ercan Çulha,

Sivas, Türkiye, culhae.55@gmail.com

Başvuru/Submitted: 15/05/2026

1. Revizyon/1st Revised: 14/07/2026

2. Revizyon/2nd Revised: 25/07/2026

Kabul/Accepted: 8/08/2026

Yayın/Online Published: 25/09/2026

Atıf/Citation: Çulha, E., Mısırlıoğlu, A., & Doğan, R. S., Acil serviste kullanılan tıbbi sarf malzemelerinin makine öğrenmesi yöntemleriyle tüketim tahmini: Bir özel hastane örneği, bmij (2026) 14 (3): 1572-1586, doi: <https://doi.org/10.15295/bmij.v14i3.2788>

Extended Abstract

Forecasting the consumption of medical consumables used in the emergency department using machine learning methods: A private hospital case

Literature

Healthcare inventory management is central to service continuity and financial sustainability because demand, lead times, and the criticality of medical supplies vary considerably. Prior research has therefore combined conventional time-series methods with machine-learning approaches for hospital logistics and emergency-department demand forecasting (Landry & Beaulieu, 2013; Saha & Ray, 2019; Vollmer et al., 2021).

Research subject

The study forecasts the monthly consumption of 10 high-volume medical consumables used in the emergency department of a private hospital and examines whether lagged emergency-department visit volume improves predictive performance.

Research purpose and importance

The purpose is to compare simple and complex forecasting algorithms under a leakage-resistant evaluation design and to determine their suitability for operational decision support. Reliable forecasts may assist aggregate budgeting and monitoring, whereas weak product-level performance may make automated replenishment unsafe.

Contribution of the article to the literature

The article focuses on an emergency-department sub-dataset rather than hospital-wide consumption, evaluates global machine-learning models against classical time-series baselines in the same hold-out period, and tests the incremental contribution of visit-volume variables through ablation analysis.

Design and method

The research is a retrospective quantitative modeling study based on aggregated institutional records. Chronological ordering was preserved throughout preprocessing, hyperparameter tuning, model selection, and independent testing.

Research type

The study is quantitative, retrospective, comparative, and predictive.

Research problems

The study asks whether monthly consumption can be forecast with temporal and lagged visit variables, which algorithm offers the best balance of cross-validation accuracy and stability, whether visit volume adds measurable information, and whether aggregate performance is sustained at the individual-item level.

Data collection method

Monthly material-consumption and emergency-department visit records for 2022–2025 were obtained from the hospital information system in aggregated form. The dataset contains 480 material-month observations for 10 products; the final six months of 2025, comprising 60 observations, were reserved for independent testing.

Quantitative analysis

Linear Regression, Ridge Regression, Artificial Neural Network, Random Forest, XGBoost, and Support Vector Regression were tuned with chronological, month-grouped TimeSeriesSplit cross-validation. Performance was assessed with R^2 , RMSE, MAE, WAPE, and MASE, and classical baselines were evaluated in the same final test period.

Research model

Monthly consumption is the dependent variable. Product identity, year, month, lag-1 emergency-department visits, and—in Variant B—the rolling mean of visits over the previous three months are the predictors. Only information available before the forecast month was used.

Research hypotheses

Because the study is comparative and predictive, no directional inferential hypotheses were specified. The pre-defined expectation tested through model comparison was that at least one candidate model would provide stable out-of-sample forecasts and that lagged visit-volume variables might add predictive information beyond product and calendar variables.

Findings and discussion

Random Forest in Variant B was selected under the pre-specified cross-validation score, whereas Linear Regression obtained the highest R^2 in the final test period. The distinction is important because the held-out test results were not used for model selection.

Findings as a result of the analysis

The selected Random Forest model achieved a cross-validation R^2 of 0.7922 (SD=0.0079), a selection score of 0.7844, and an independent test R^2 of 0.887. Its test WAPE was 22.52% and MASE was 1.153. Eight of the 10 products nevertheless had negative product-level test R^2 values.

Hypothesis test results

No conventional statistical hypothesis tests were conducted. Predictive expectations were evaluated through chronological cross-validation, an untouched hold-out test set, and ablation analysis. Visit-volume variables produced no measurable incremental gain for the selected Random Forest model, although gains appeared in the linear models.

Discussing the findings with the literature

The results support literature emphasizing the value of data-driven healthcare logistics while also showing that strong aggregate accuracy can mask weak item-level forecasts. The competitiveness of simple baselines and linear models is consistent with evidence that model complexity does not guarantee superior forecasting performance.

Conclusion, recommendation and limitations

The model is suitable only as an auxiliary, human-supervised tool for aggregate monitoring and budgeting. Product-level ordering should continue to incorporate inventory position, lead time, clinical requirements, and expert review.

Results of the article

The study identifies Variant B–Random Forest as the cross-validation-based choice but finds limited incremental value from visit volume in that model and weak item-level generalization for most products.

Suggestions based on results

Future research should use longer time series, multiple rolling-origin test windows, multiple hospitals, and richer variables such as triage level, diagnosis group, procedure volume, stock level, lead time, unit cost, and seasonal disease indicators.

Limitations of the article

The evidence comes from one private hospital, 48 months, 10 products, three cross-validation folds, and one six-month hold-out window. Monthly aggregation and the absence of clinical, stock, cost, and procurement variables limit causal interpretation and external validity.

Giriş

Sağlık sektörü, küresel ölçekte hem en büyük hem de en hızlı büyüyen alanlardan biri olarak kabul edilmektedir. Türkiye’de sağlık harcamalarının son yıllarda önemli ölçüde arttığı görülmektedir. Türkiye İstatistik Kurumu (2024) verilerine göre 2022 yılında toplam sağlık harcamaları 606,8 milyar TL olarak gerçekleşmiş ve bu tutar Gayri Safi Yurt İçi Hasıla’nın (GSYH) %4’ünü oluşturmuştur. 2023 yılında ise sağlık harcamaları 1,24 trilyon TL seviyesine yükselmiş (GSYH’nin %4,7’si) ve kişi başına sağlık harcaması 14.582 TL’ye ulaşmıştır. Bu artış, sağlık sisteminde maliyetlerin kontrolü ve kaynak kullanımının etkinliği konusunu daha da önemli hâle getirmektedir. Nitekim artan sağlık harcamaları, politika yapıcılar açısından yüksek kaliteli sağlık hizmeti sunulurken kaynakların nasıl daha verimli kullanılabileceği sorusunu gündeme getirmektedir (Ayer vd., 2023). Bu kapsamda tıbbi malzeme, ilaç ve ekipmanların planlanması, izlenmesi ve kontrolünü içeren envanter yönetimi, sağlık hizmetlerinin kesintisiz sürdürülebilmesi açısından kritik öneme sahiptir (Khokhar, 2023; Landry ve Beaulieu, 2013; Saha ve Ray, 2019).

Acil servis başvurularının ve kullanılan tıbbi malzemelerin tahmin edilmesi, envanter planlamasının geliştirilmesine, hasta akışının düzenlenmesine ve olası yoğunluklara karşı hazırlık yapılmasına katkı sağlayabilir (Asheim vd., 2019; Jones vd., 2008; Tuominen vd., 2022; Vollmer vd., 2021). Bu kapsamda acil servislerde envanter yönetiminin önemi daha da belirginleşmektedir. Acil servislerde hizmetin ertelenmesi veya malzeme eksikliğine bağlı gecikme oluşması hasta güvenliği açısından risk doğurabilir. Bu nedenle geçmiş tüketim verilerinin yanı sıra hasta başvuru sayısı gibi hizmet hacmi göstergelerinin tahmin modellerinde sağlanabileceği ek katkı da değerlendirilmelidir.

Yapay zekâ ve makine öğrenmesi yöntemlerinin envanter yönetimine entegrasyonu, geniş veri kümelerinin analizini ve tahmin süreçlerinin otomasyonunu destekleyebilir (Ali vd., 2025). Tıbbi malzeme tüketiminin tahmini de stok planlaması ve kontrol süreçlerinin daha sistematik yürütülmesine katkı sağlayabilir. Bununla birlikte hizmet gecikmeleri ve hizmet kalitesi üzerindeki olası etkiler, kurumun veri yapısı ve uygulama koşulları çerçevesinde ayrıca değerlendirilmelidir (Sayın, 2016).

Bu çalışmanın amacı, özel bir hastanenin acil servisinde kullanılan yüksek tüketimli tıbbi sarf malzemelerinin aylık tüketim miktarlarını makine öğrenmesi yöntemleriyle tahmin etmek ve acil servis başvuru sayısının tahmin sürecindeki rolünü değerlendirmektir. Çalışmanın literatüre katkısı üç noktada özetlenebilir. Birincisi, hastane geneli tüketim tahminlerinden farklı olarak acil servis özelinde bir alt veri setine odaklanmakta ve birim düzeyinde stok yönetimi kararlarını ele almaktadır. İkincisi, yalnızca model doğruluğunu değil; çapraz doğrulama istikrarı ve kurumsal bilgi sistemlerine entegre edilebilirlik gibi yönetsel uygulanabilirlik ölçütlerini de dikkate almaktadır. Üçüncüsü, sağlık kurumlarında rutin olarak izlenen hasta başvuru yoğunluğunun tıbbi sarf malzeme tüketim tahminindeki olası katkısını sınamakta ve bu değişkenin modele sağladığı artımlı bilginin kullanılan algoritmaya göre değişebildiğini göstermektedir.

Bu çerçevede çalışma aşağıdaki araştırma sorularına yanıt aramaktadır: (1) Acil serviste kullanılan yüksek tüketimli tıbbi sarf malzemelerinin aylık tüketim miktarları, acil servis hasta başvuru sayısı ve zamansal değişkenler kullanılarak makine öğrenmesi yöntemleriyle güvenilir biçimde tahmin edilebilir mi? (2) Karşılaştırılan algoritmalar arasında hangi model, çapraz doğrulama performansı ve katlar arası istikrar birlikte değerlendirildiğinde en uygun sonucu vermektedir? (3) Hasta başvuru sayısının tahmin gücü, malzemeler arası toplam tüketim düzeyinde mi yoksa tek bir ürünün ay bazlı tüketim düzeyinde mi geçerlidir?

Kavramsal çerçeve ve literatür

Sağlık kurumlarında değişken talep yapısı ve kritik malzemelerin sürekli hazır bulundurulması gerekliliği, stok planlamasını karmaşıklştırmaktadır. Bu nedenle tüketim eğilimlerini geçmiş verilerden öğrenebilen tahminleme yöntemleri, tıbbi malzeme yönetiminde giderek daha fazla kullanılmaktadır (Ali vd., 2025; Khokhar, 2023; Langabeer ve Helton, 2021).

Tıbbi sarf malzeme talebi düzenli, değişken, yavaş hareketli veya aralıklı bir yapı gösterebilmektedir. Sıfır talep dönemlerinin sık olduğu ürünlerde geleneksel yöntemler yetersiz kalabildiğinden Croston ve türevleri gibi aralıklı talep yöntemleri kullanılmaktadır. Liu (2020), kısa yaşam döngülü tıbbi sarf malzemelerinde talep yapısına uygun toplulaştırma, model seçimi ve hata ölçütlerinin önemini ortaya koymuştur. Bu çalışmadaki ürünler aylık düzeyde sıfır tüketim içermediği için Croston türü modeller uygulanmamış, ilgili literatür yöntemsel kapsamın açıklanmasında kullanılmıştır.

Çoklu ürün tahminlerinde her ürün için ayrı model kuran yerel yaklaşımlar ile bütün ürünleri ortak bir modelde değerlendiren küresel yaklaşımlar kullanılmaktadır. ETS, ARIMA ve ARIMAX ürün bazlı yerel modellerken, ürün bilgisini ortak bir veri yapısında kullanan makine öğrenmesi modelleri küresel

yaklaşım niteliğindedir. Montero-Manso ve Hyndman (2021), küresel modellerin kısa ve heterojen serilerde de rekabetçi sonuçlar verebildiğini göstermektedir. Bununla birlikte toplu performansın yüksek olması her ürünün aylık değişiminin başarılı biçimde tahmin edildiği anlamına gelmediğinden ürün bazlı hata ölçütlerinin de raporlanması gerekmektedir.

Makine öğrenmesi yöntemleri, geçmiş verilerdeki doğrusal ve doğrusal olmayan ilişkilerden yararlanarak tahmin üretmektedir. Bu çalışmada Lineer Regresyon, Ridge Regresyon, Random Forest, XGBoost, SVR ve Yapay Sinir Ağı modelleri karşılaştırılmıştır. Doğrusal modeller yorumlanabilirlik sağlarken, ağaç tabanlı modeller, SVR ve yapay sinir ağları daha karmaşık ilişkileri öğrenebilmektedir. Ancak model başarısı veri miktarı, değişken yapısı ve hiperparametre seçimlerinden etkilendiğinden karmaşık modellerin her durumda daha başarılı olması beklenmemelidir (Chen ve Guestrin, 2016; Goodfellow vd., 2016; James vd., 2013).

Sağlık alanındaki çalışmalar da model performansının veri yapısına göre değiştiğini göstermektedir. Altın vd. (2023), özel bir hastanede tıbbi atık miktarının tahmininde derin öğrenmenin çekirdek tabanlı SVM'den daha iyi sonuç verdiğini bildirmiştir. Vollmer vd. (2021) ise acil servis başvurularında doğrusal ve geleneksel zaman serisi modellerinin bazı durumlarda karmaşık makine öğrenmesi yöntemleri kadar başarılı olabildiğini göstermiştir. Bu bulgular, makine öğrenmesi modellerinin basit referanslar, ETS ve ARIMA ailesiyle karşılaştırılması gerektiğini desteklemektedir.

Mevcut literatürde acil servis başvurularının tahmini yaygın olmakla birlikte, belirli tıbbi sarf malzemelerinin çoklu ürün yapısında tahmin edilmesi, hasta başvuru sayısının artımlı katkısının sınanması ve küresel makine öğrenmesi modellerinin ürün bazlı ETS, ARIMA ve ARIMAX modelleriyle aynı test döneminde karşılaştırılması sınırlıdır. Bu çalışma söz konusu yöntemleri yüksek tüketimli on ürün üzerinde karşılaştırmakta ve sonuçları toplu ölçütlerin yanında ürün bazlı R^2 ve WAPE değerleriyle değerlendirmektedir. Ürün bazındaki sınırlı performans nedeniyle çalışmanın katkısı otomatik sipariş kararlarından ziyade yöntem karşılaştırması ve insan denetimli toplu karar desteğiyle sınırlandırılmıştır.

Yöntem

Araştırma tasarımı

Bu araştırma, retrospektif kurumsal kayıt verilerine dayalı, nicel ve tahmine yönelik bir modelleme çalışması olarak tasarlanmıştır. Çalışmada, acil servis biriminde kullanılan yüksek tüketimli tıbbi sarf malzemelerinin aylık tüketim miktarları makine öğrenmesi algoritmaları ile tahmin edilmiştir. Araştırma kapsamında farklı algoritmaların tahmin performansları karşılaştırmalı olarak değerlendirilmiştir.

Çalışma, doktora tezi kapsamında kullanılan genel hastane tıbbi malzeme tüketim veri setinden türetilmiş olup yalnızca acil servis biriminde kullanılan tıbbi malzemelere odaklanan bir alt örneklem analizi olarak yapılandırılmıştır. Bu yönüyle çalışma, hastane geneli tüketim tahmininden farklı olarak acil servis hizmet yoğunluğu ile tıbbi malzeme tüketimi arasındaki ilişkiye odaklanmaktadır.

Bu çalışmada kullanılan veriler, ilgili özel hastanenin kurumsal bilgi sisteminden hasta veya çalışan kimliğini belirlemeye imkân verecek herhangi bir bilgi içermeyecek şekilde, malzeme ve ay düzeyinde toplulaştırılmış olarak temin edilmiştir. Çalışma, ilgili özel hastane yönetimi tarafından 10.01.2025 tarihli ve 2025/112 sayılı yazı ile verilen araştırma izni kapsamında yürütülmüştür. Araştırmacılar çalışma süresince hasta veya çalışan kimliğini belirleyebilecek herhangi bir veriye erişmemiştir. Araştırmada bireysel hasta verisi, kişisel sağlık verisi, anket, görüşme veya klinik müdahale yer almadığından etik kurul onayı gerekmemektedir.

Veri seti ve değişkenler

Araştırmada 2022–2025 yılları arasındaki aylık kayıtlar kullanılmıştır. Veri seti, acil serviste kullanılan ve toplam tüketim miktarı bakımından öne çıkan 10 tıbbi sarf malzemeye ait 480 gözlemden oluşmaktadır. Her gözlem, belirli bir malzemenin belirli bir yıl ve aya ait tüketim miktarını göstermektedir.

Analizde kullanılan değişkenler Tablo 1’de sunulmuştur.

Tablo 1: Araştırmada Kullanılan Değişkenler

Değişken	Rolü	Açıklama
Tüketim miktarı	Bağımlı değişken	İlgili malzemenin aylık tüketim miktarı
Malzeme adı	Bağımsız değişken	Tıbbi sarf malzemenin adı
Yıl	Bağımsız değişken	Gözlemin ait olduğu yıl
Ay	Bağımsız değişken	Gözlemin ait olduğu ay
Gecikmeli acil servis başvuru sayısı	Bağımsız değişken	Tahmin ayından önce gerçekleşen hasta başvuru bilgisi; Varyant A'da lag-1, Varyant B'de lag-1 ve önceki üç ayın hareketli ortalaması

Kaynak: Yazarlar tarafından üretilmiştir.

Veri ön işleme

Veri ön işleme sürecinde öncelikle yıl ve ay değişkenlerinden tarih değişkeni oluşturulmuş, veri seti kronolojik olarak sıralanmıştır. Zaman serisi niteliği taşıyan verilerde rastgele sıralama veya rastgele eğitim-test ayrımı veri sızıntısına neden olabileceğinden, modelleme sürecinde zamansal sıra korunmuştur.

Aykırı değer kontrolünde Medyan Mutlak Sapma yöntemi kullanılmıştır. Bu yöntem, özellikle küçük veya çarpık dağılımlı veri setlerinde standart sapmaya dayalı yöntemlere göre daha gülbüz bir yaklaşım sunmaktadır (Leys vd., 2013). Aykırı değer kontrolü yalnızca bağımsız değişken olan acil servis başvuru sayısı üzerinde uygulanmış, bağımlı değişken olan tüketim miktarı kırılmamıştır. Böylece modelin gerçek tüketim davranışını öğrenmesi amaçlanmıştır. Eğitim setinde acil servis başvuru sayısı için üst sınır 8.921,80 hasta/ay olarak belirlenmiş ve bu sınırın üzerinde kalan 10 gözlem üst sınıra kırılmıştır. Ay ve yıl değişkenlerinde aykırı değer müdahalesi gerektiren bir durum saptanmamıştır; final test setinde ise herhangi bir kırpma işlemi yapılmamıştır.

Sayısal değişkenler, ortak bir ölçekte değerlendirilebilmesi ve modelin sayısal kararlılığının artırılması amacıyla merkezleme ve ölçeklendirme (z-puanı standardizasyonu / StandardScaler) yöntemi ile normalize edilmiştir. Kategorik değişkenler ise analize uygun hâle getirilmek üzere, her bir kategoriye ayrı bir gösterge değişkenine dönüştüren kukla değişken (one-hot encoding) yaklaşımı ile yeniden yapılandırılmıştır (Kuhn ve Johnson, 2013). Bu kapsamda, kategorik değişken olan malzeme adı da one-hot encoding yöntemi kullanılarak sayısallaştırılmış ve modele dâhil edilmiştir. Ayrıca doğrusal modeller, yapay sinir ağları ve SVR algoritmaları için sayısal değişkenler StandardScaler yöntemi ile standartlaştırılarak modele uygun hâle getirilmiştir.

Veri sızıntısı kontrolü ve gecikmeli (lag) değişkenler

Acil servise başvuran hasta sayısının ilgili aya ait gerçekleşmiş değeri, tahmin anında henüz bilinmeyebileceğinden ileriye dönük tahmin sürecinde veri sızıntısına yol açabilmektedir. Bu nedenle analizde hasta başvuru sayısı yalnızca tahmin anında erişilebilir olan geçmiş dönem bilgileri kullanılarak yeniden yapılandırılmıştır. Bu kapsamda iki alternatif özellik kümesi karşılaştırmalı olarak test edilmiştir: (i) Varyant A, bir önceki ayın gerçekleşmiş hasta başvuru sayısını (lag-1) kullanmaktadır; (ii) Varyant B, lag-1 değişkenine ek olarak önceki üç ayın hareketli ortalamasını (roll3) içermektedir. Her iki değişken kümesi de yalnızca geçmişte gerçekleşmiş verilerden türetildiğinden, aylık operasyonel karar destek yaklaşımıyla uyumludur. Aykırı değer kırpması (Medyan Mutlak Sapma), gecikmeli değişkenler türetilmeden önce ham hasta sayısı serisine uygulanmış, kırılmış değerler daha sonra lag ve hareketli ortalama hesaplamalarına yansıtılmıştır.

Eğitim ve test seti ayrımı

Veri setinde 2025 yılı Temmuz–Aralık dönemine ait son 6 ay bağımsız final test seti olarak korunmuştur. Ancak gecikmeli değişkenlerin türetilmesi serinin başında bir ısınma (warm-up) dönemi gerektirdiğinden, Varyant A için ilk ay, Varyant B için ilk üç ay eğitim setinden çıkarılmıştır. Buna göre eğitim ve çapraz doğrulama sürecinde Varyant A'da 410, Varyant B'de 390 gözlem kullanılmış; final test seti her iki varyantta da 60 gözlem olarak sabit kalmıştır. Final test setindeki tahminler için kullanılan gecikmeli hasta sayısı değerleri de her durumda ilgili ayın kendisinden önceki, dolayısıyla tahmin anında zaten gerçekleşmiş aylara aittir.

Modelin etkinliğinin tarafsız biçimde değerlendirilebilmesi için model, kurulum ve ayarlama aşamalarında kullanılmamış veriler üzerinde sınanmalıdır. Bu kapsamda modelin oluşturulduğu veri kümesi eğitim ve doğrulama, performansının değerlendirildiği veri kümesi ise test seti olarak adlandırılmaktadır (Kuhn ve Johnson, 2013). Veri setinin panel yapısı (10 malzeme × 48 ay) dikkate alınarak çapraz doğrulamada TimeSeriesSplit yöntemi satır düzeyinde değil, ay düzeyinde gruplanmış

olarak uygulanmıştır: bölme noktaları yalnızca ay sınırlarında oluşturulmuş, aynı aya ait 10 malzeme satırı hiçbir zaman farklı katlara dağılmamıştır. Bu yaklaşım hem klasik K-Fold yönteminde ortaya çıkabilecek zamansal karışmayı hem de satır bazlı bölmenin yol açabileceği yapay gözlem bağımsızlığı varsayımını ortadan kaldırmaktadır. Üç kat uygulanmış; yalnızca kat ortalaması değil, katlar arası R^2 standart sapması (CV R^2 Std) da raporlanarak model istikrarı değerlendirilmiştir. Ön işleme adımları (one-hot encoding ve StandardScaler) her çapraz doğrulama katında ve nihai eğitimde yalnızca ilgili eğitim verisine uyarlanmış, doğrulama/test verisine yalnızca dönüştürme uygulanmıştır; bu işlem veri sızıntısının önlenmesi açısından önemlidir.

Kullanılan algoritmalar ve performans metrikleri

Çalışmada Ridge Regresyon, Lineer Regresyon, Yapay Sinir Ağı, Random Forest, XGBoost ve Destek Vektör Regresyonu algoritmaları karşılaştırmalı olarak değerlendirilmiştir. Ridge Regresyon, doğrusal modellerde ortaya çıkabilecek çoklu doğrusal bağlantı sorununu ve aşırı uyum riskini azaltmaya yönelik cezalandırmalı bir tahminleme yaklaşımı sunması nedeniyle analiz kapsamına alınmıştır (Hoerl ve Kennard, 1970). Random Forest algoritması, çok sayıda karar ağacının birleşik yapısından yararlanarak doğrusal olmayan ilişkileri ve değişkenler arasındaki karmaşık etkileşimleri modelleyebilmesi bakımından sağlık hizmetlerinde değişkenlik gösteren tüketim örüntülerinin tahmin edilmesine uygun bir yöntem olarak değerlendirilmiştir (Géron, 2019). XGBoost ise ağaç tabanlı gradyan artırma yaklaşımına dayanması, karmaşık ve çok boyutlu veri yapılarında ölçeklenebilirlik sağlaması ve yüksek tahmin başarısı sunabilmesi nedeniyle tercih edilmiştir (Chen ve Guestrin, 2016). Bunun yanında Destek Vektör Regresyonu ve Yapay Sinir Ağı modelleri de doğrusal olmayan veri örüntülerini yakalayabilme potansiyelleri nedeniyle karşılaştırmalı modelleme sürecine dâhil edilmiştir (Géron, 2019, pp. 157-171; Goodfellow vd., 2016).

Model hiperparametrelerinin belirlenmesinde sistematik bir optimizasyon süreci uygulanmıştır. GridSearchCV yöntemi yalnızca eğitim verisi üzerinde, ay düzeyinde gruplanmış üç katlı çapraz doğrulama kullanılarak gerçekleştirilmiş; final test seti hiperparametre seçimi ve model belirleme sürecine dâhil edilmemiştir.

Ridge Regresyon için $\alpha \in \{0,1; 1; 10; 100\}$ değerleri sınanmış ve $\alpha=1$ seçilmiştir. Random Forest için ağaç sayısı $\in \{300, 600\}$ ve maksimum derinlik $\in \{\text{sınırsız}, 10\}$ değerleri değerlendirilmiş; 300 ağaç ve sınırsız maksimum derinlik en uygun parametreler olarak belirlenmiştir. XGBoost için ağaç sayısı $\in \{300, 600\}$, maksimum derinlik $\in \{3, 6\}$ ve öğrenme oranı $\in \{0,05; 0,1\}$ değerleri sınanmış; 300 ağaç, maksimum derinlik 3 ve öğrenme oranı 0,05 seçilmiştir. Yapay Sinir Ağı için gizli katman yapısı $\in \{(32), (64, 32)\}$ ve L2 düzenleme katsayısı $\in \{0,0001; 0,001; 0,01\}$ değerleri değerlendirilmiş; (64, 32) katman yapısı ve 0,001 düzenleme katsayısı kullanılmıştır. SVR için C $\in \{1, 10, 100\}$, epsilon $\in \{0,1; 1\}$ ve çekirdek türü $\in \{\text{radyal tabanlı}, \text{doğrusal}\}$ değerleri sınanmış; C=100, epsilon=0,1 ve doğrusal çekirdek seçilmiştir. Lineer Regresyon modeli ayarlanabilir bir hiperparametre içermediğinden varsayılan yapı ile uygulanmıştır. Rastgelelik içeren modellerde random_state=42 olarak sabitlenmiştir.

Analizler Python 3.12 ortamında gerçekleştirilmiştir. Makine öğrenmesi modellerinin oluşturulması ve değerlendirilmesinde scikit-learn 1.8, XGBoost modelinde xgboost 3.3, ETS ve ARIMA modellerinde statsmodels 0.14, otomatik ARIMA model seçimi süreçlerinde ise pmdarima 2.1 kütüphanelerinden yararlanılmıştır.

Model seçim protokolü ve hasta sayısının artımlı katkısının sınanması

En iyi model, yalnızca çapraz doğrulama sonuçlarına dayanan ve önceden tanımlanan bir kurala göre seçilmiştir. Her modelin üç fold'daki ortalama R^2 değerinden katlar arası standart sapma çıkarılarak "Seçim Skoru" (CV R^2 Ortalaması – CV R^2 Standart Sapması) hesaplanmış; en yüksek skora sahip model en iyi model olarak belirlenmiştir. Bu ölçüt, doğruluk ile katlar arası istikrarı birlikte değerlendiren muhafazakâr bir seçim ölçütüdür. Final test seti, en iyi model çapraz doğrulama sonuçlarına göre belirlendikten sonra yalnızca bağımsız doğrulama amacıyla bir kez kullanılmış; model karşılaştırması veya seçimi test sonuçlarına dayandırılmamıştır. Hasta başvuru sayısı değişkenlerinin (lag-1 ve roll3) modele artımlı katkısını sınamak amacıyla aynı hiperparametre optimizasyonu ve çapraz doğrulama protokolü, hasta sayısı değişkenlerinin çıkarıldığı özellik kümesiyle (yalnızca malzeme adı, ay ve yıl) tekrarlanmış; iki özellik kümesinin çapraz doğrulama R^2 değerleri arasındaki fark değişken çıkarma (ablation) analizi olarak raporlanmıştır. Ayrıca hasta sayısının lag-1 değerini dışsal değişken olarak kullanan ARIMAX modeli klasik zaman serisi karşılaştırmasına eklenmiştir. Aynı hiperparametre optimizasyonu, ay düzeyinde gruplanmış çapraz doğrulama ve değişken çıkarma protokolü Varyant A ile Varyant B için ayrı ayrı uygulanmış; genel en iyi model, iki varyanttaki toplam 12 model sonucu arasından yalnızca Seçim Skoruna göre belirlenmiştir.

Model performansının değerlendirilmesinde R^2 temel performans ölçütü olarak kullanılmıştır. R^2 , modelin tüketim miktarındaki varyansı ne ölçüde açıkladığını göstermektedir. Buna ek olarak RMSE ve MAE değerleri raporlanmıştır. RMSE büyük hatalara daha duyarlı olduğundan yüksek sapmaları öne çıkarırken, MAE ortalama mutlak hata düzeyini göstermektedir (Hyndman ve Koehler, 2006; Kuhn ve Johnson, 2013). En iyi model seçimi, yalnızca çapraz doğrulama ortalaması ile katlar arası istikrarı birlikte dikkate alan Seçim Skoruna göre gerçekleştirilmiştir. İncelenen 10 malzemenin tüketim hacimleri önemli ölçüde farklılaştığından WAPE (Weighted Absolute Percentage Error), toplam mutlak hatanın toplam gerçekleşen tüketime oranını göstermek amacıyla hem çapraz doğrulama hem de final test aşamasında toplam ve ürün bazlı düzeyde raporlanmıştır. Hyndman ve Koehler (2006) tarafından önerilen ölçekten bağımsız MASE (Mean Absolute Scaled Error) de hesaplanmış; bu ölçüt, her modelin hatasını ilgili malzemenin eğitim setindeki naive (bir önceki ay) tahmininin ortalama mutlak hatasına oranlayarak ürünler arası karşılaştırmaya imkân sağlamıştır.

Bulgular

Veri setinin genel yapısı

Analiz veri seti 10 tıbbi malzemeye ait 480 aylık gözlemden oluşmaktadır. İncelenen dönemde bu ürünlerin toplam tüketim miktarı 689.873 birimdir. Aylık acil servis başvuru sayısı 1.904 ile 9.152 arasında değişmektedir.

Tablo 2: İncelenen Malzemelerin Toplam Tüketim Miktarına Göre Sıralaması

Malzeme adı	Toplam tüketim miktarı
Enjektör 10 ml Üç Parçalı Yeşil 21G	186.898
Serum Seti İğneli	167.077
EKG Elektrot Yetişkin	142.985
İntravenöz Kanül Portlu Pembe No:20	76.654
İntravenöz Kanül Portlu Mavi No:22	53.833
İntravenöz Kanül Portlu Sarı No:24	24.004
İntravenöz Kanül Portlu Mor No:26	14.419
Nebülizer Set Maske Yetişkin	9.742
Enjektör İnsülin Tek Kullanımlık 26G 100 IU 1 ml	7.703
Nebülizer Set Maske Pediatrik	6.558

Kaynak: Yazarlar tarafından üretilmiştir.

Acil servis başvuru sayısındaki aykırı değerlerin değerlendirilmesi

Eğitim setinde acil servis başvuru sayısı için medyan 4.096, medyan mutlak sapma 930 olarak hesaplanmıştır. Z-skor eşik değeri $\pm 3,5$ kabul edildiğinde üst sınır 8.921,80 hasta/ay olarak belirlenmiştir. Eğitim setinde bu sınırın üzerinde kalan 10 gözlem kırpılmıştır. Bu işlem bağımsız değişkendeki aşırı uç değerlerin model eğitimini bozmasını engellemeyi amaçlamış, tüketim miktarı üzerinde herhangi bir kırpma yapılmamıştır.

Model performanslarının karşılaştırılması

Modellerin hiperparametre optimizasyonu sonrasındaki çapraz doğrulama sonuçları ve model seçim skorları, Varyant A (Tablo 3a) ve Varyant B (Tablo 3b) için ayrı ayrı sunulmuştur. Her iki varyant aynı GridSearchCV protokolü, aynı ay düzeyinde gruplanmış çapraz doğrulama bölmeleri ve aynı Seçim Skoru ölçütüyle değerlendirilmiştir. En iyi model, iki tablodaki toplam 12 model-varyant kombinasyonu arasından en yüksek Seçim Skoruna sahip model olarak belirlenmiştir. Ayrıca modellerin gerçek katkısını değerlendirmek amacıyla altı baz/klasik zaman serisi modeli karşılaştırmaya eklenmiş ve hasta sayısı değişkenlerinin artımlı katkısı Varyant B için değişken çıkarma analiziyle incelenmiştir. Varyant A'ya ilişkin değişken çıkarma sonuçları ise metinde ayrıca değerlendirilmiştir.

Tablo 3a: Varyant A (Lag-1 Hasta Sayısı) – Hiperparametre Optimizasyonu Sonrası Çapraz Doğrulama ve Seçim Skoru

Model	CV R ² Ort.	CV R ² Std	Seçim Skoru (Ort.-Std)	Test R ² *	Test MASE*
Lineer Regresyon	0,7890	0,0492	0,7398	0,918	1,238
Ridge Regresyon	0,7886	0,0497	0,7390	0,913	1,244
Yapay Sinir Ağı (ANN)	0,7599	0,0322	0,7277	0,917	1,197
Random Forest	0,7340	0,0978	0,6363	0,860	1,354
XGBoost	0,7059	0,1405	0,5654	0,839	1,332
SVR	0,6014	0,1730	0,4283	0,857	1,308

Kaynak: Yazarlar tarafından üretilmiştir. **Not:** Test sonuçları yalnızca bağımsız doğrulama amacıyla sunulmuş, model seçiminde kullanılmamıştır. Sıralama Seçim Skoruna göre yapılmıştır.

Tablo 3b: Varyant B (Lag-1 + 3 Aylık Hareketli Ortalama) – Hiperparametre Optimizasyonu Sonrası Çapraz Doğrulama ve Seçim Skoru

Model	CV R ² Ort.	CV R ² Std	Seçim Skoru (Ort.-Std)	Test R ² *	Test WAPE (%) *	Test MASE*
Random Forest	0,7922	0,0079	0,7844	0,887	22,52	1,153
XGBoost	0,7850	0,0113	0,7737	0,901	22,53	1,299
Yapay Sinir Ağı (ANN)	0,7946	0,0306	0,7640	0,885	27,20	2,177
Ridge Regresyon	0,7372	0,0380	0,6992	0,917	21,01	1,131
Lineer Regresyon	0,7239	0,0296	0,6943	0,921	20,45	1,063
SVR	0,6404	0,1276	0,5128	0,850	26,09	1,183

Kaynak: Yazarlar tarafından üretilmiştir. **Not:** Test sonuçları yalnızca bağımsız doğrulama amacıyla sunulmuş, model seçiminde kullanılmamıştır. Sıralama Seçim Skoruna göre yapılmıştır.

Tablo 3b’de tanımlanan Seçim Skoruna (CV R² Ortalaması – CV R² Std) göre en yüksek değer Random Forest modelinde elde edilmiştir (0,7844). Bu modeli XGBoost (0,7737) ve ANN (0,7640) izlemiştir. Random Forest’in çapraz doğrulama R² değeri 0,7922 ve standart sapması 0,0079 olarak bulunmuştur. ANN daha yüksek bir ortalama çapraz doğrulama R² değerine ulaşmasına rağmen (0,7946), standart sapmasının daha yüksek olması (0,0306) nedeniyle Seçim Skoru Random Forest’in altında kalmıştır. Model seçimi yalnızca çapraz doğrulama sonuçlarına göre gerçekleştirilmiş ve final test seti bu aşamada kullanılmamıştır. Seçilen Random Forest modelinin final test performansı R²=0,887, WAPE=%22,52 ve MASE=1,153 olarak hesaplanmıştır. Lineer Regresyon final testte R²=0,921 değerine ulaşmış olmakla birlikte, final test sonuçları model seçim ölçütü olarak kullanılmamıştır.

Varyant A, aynı hiperparametre optimizasyonu ve ay düzeyinde gruplanmış çapraz doğrulama protokolüyle değerlendirilmiştir. Bu varyantta en yüksek Seçim Skoru Lineer Regresyon modelinde elde edilmiş (0,7398), ancak bu değer Varyant B’deki Random Forest modelinin Seçim Skorundan (0,7844) düşük kalmıştır. Bu nedenle iki varyanttaki toplam 12 model-varyant kombinasyonu arasında en iyi model Varyant B–Random Forest olarak belirlenmiştir. Varyant A’nın değişken çıkarma analizinde hasta sayısı değişkenlerinin çıkarılması, Random Forest modelinde çapraz doğrulama R² değerini 0,054, XGBoost modelinde ise 0,031 azaltmıştır.

SVR modeli en düşük Seçim Skoruna sahiptir (0,5128) ve final testte de en zayıf performansı göstermiştir (R²=0,850). Sistemik hiperparametre taraması sonunda doğrusal çekirdek ve C=100 seçilmiş olmasına rağmen bu model veri yapısına en uygun seçenek olmamıştır. Final test MASE değerleri küçükten büyüğe Lineer Regresyon için 1,063, Ridge Regresyon için 1,131, Random Forest için 1,153, SVR için 1,183, XGBoost için 1,299 ve ANN için 2,177 olarak hesaplanmıştır. Naive baz modelin MASE değeri (0,989) tüm makine öğrenmesi modellerinden daha düşüktür. Bu bulgu, ölçekten bağımsız MASE ölçütüne göre basit bir süreklilik yaklaşımının rekabetçi kaldığını göstermektedir.

Baz (naive) tahmin modelleriyle karşılaştırma

Karmaşık modellerin gerçek katkısını değerlendirebilmek amacıyla altı baz/klasik zaman serisi modeli final test döneminde ayrıca hesaplanmıştır: (i) Naive (t-1), (ii) Mevsimsel Naive (t-12), (iii) Hareketli Ortalama (3 ay), (iv) ETS (üstel düzeltme), (v) otomatik ARIMA ve (vi) hasta sayısının lag-1 değerini dışsal değişken olarak kullanan otomatik ARIMAX. Sonuçlar Tablo 3c’de sunulmuştur.

Tablo 3c: Baz (Naive/Klasik Zaman Serisi) Tahmin Modellerinin Nihai Test Performansı

Baz Model	Test R ²	Test RMSE	Test WAPE (%)	Test sMAPE (%)	MASE
Naive (t-1)	0,883	550,00	22,32	42,14	0,989
Mevsimsel Naive (t-12)	0,803	714,94	31,21	56,77	1,440
Hareketli Ortalama (3 ay)	0,891	531,72	21,70	46,46	1,049
ETS (Üstel Düzeltme)	0,824	676,54	28,66	48,42	1,227
ARIMA (otomatik)	0,858	606,11	25,61	45,88	1,129
ARIMAX (otomatik, hasta sayısı dışsal)	0,834	656,86	27,46	46,42	1,168

Kaynak: Yazarlar tarafından üretilmiştir.

Tablo 3c incelendiğinde, yalnızca bir önceki ayın değerini tekrar eden Naive modelin $R^2=0,883$ ve $WAPE=\%22,32$ değerlerine ulaştığı görülmektedir. Bu performans, karşılaştırılan makine öğrenmesi modellerinin çoğuna yakındır ve bazı ölçütlerde daha iyidir. ETS, ARIMA ve ARIMAX modelleri ise sırasıyla $R^2=0,824$, $R^2=0,858$ ve $R^2=0,834$ değerleriyle Naive ve hareketli ortalama baz modellerinin gerisinde kalmıştır. Hasta sayısını dışsal değişken olarak kullanmasına rağmen ARIMAX modelinin ARIMA'dan daha iyi sonuç vermemesi, hasta sayısı bilgisinin klasik zaman serisi çerçevesinde ek açıklama gücü sağlamadığını göstermektedir. Bu bulgu, Tablo 3d'deki değişken çıkarma analiziyle de uyumludur. Random Forest modeli (Varyant B, $R^2=0,887$; $WAPE=\%22,52$), Naive baz modele göre WAPE bakımından 0,20 yüzde puan daha kötü sonuç vermiştir. Sonuçlar, hasta başvuru sayısının bazı modellerde ölçülebilir katkı sağladığını, ağaç tabanlı modellerde bu katkının belirsiz kaldığını ve model performansının tek bir ölçüte dayanılarak değerlendirilmemesi gerektiğini göstermektedir.

Hasta sayısının artımlı katkısı: Değişken çıkarma analizi

Hasta başvuru sayısı değişkenlerinin (lag-1 ve roll3) modele gerçekte ne kadar katkı sağladığını doğrudan sınamak amacıyla, her algoritma aynı hiperparametre optimizasyonu ve ay-gruplu çapraz doğrulama protokolüyle, hasta sayısı değişkenleri çıkarılmış bir özellik kümesiyle (yalnızca malzeme adı, ay, yıl) yeniden eğitilmiştir. Sonuçlar Tablo 3d'de sunulmaktadır.

Tablo 3d: Hasta Sayısı Değişkenlerinin Artımlı Katkısı (Ablation Analizi, Çapraz Doğrulama)

Model	CV R ² (Tam Model)	CV R ² (Hasta Sayısı Hariç)	Fark
Random Forest	0,7922	0,8083	-0,0160
XGBoost	0,7850	0,7937	-0,0087
Yapay Sinir Ağı (ANN)	0,7946	-0,1525	+0,9471
Ridge Regresyon	0,7372	0,4220	+0,3152
Lineer Regresyon	0,7239	0,3248	+0,3991
SVR	0,6404	0,6755	-0,0351

Kaynak: Yazarlar tarafından üretilmiştir.

Tablo 3d, hasta sayısı değişkenlerinin katkısının model ailesine göre farklılaştığını göstermektedir. Ağaç tabanlı modellerde (Random Forest ve XGBoost) hasta sayısı değişkenlerinin çıkarılması çapraz doğrulama R^2 değerini sırasıyla 0,016 ve 0,009 artırmıştır. Buna göre seçilen Random Forest modelinde hasta başvuru sayısı ölçülebilir bir artımlı katkı sağlamamıştır. Buna karşılık Lineer Regresyon ve Ridge Regresyon modellerinde hasta sayısı değişkenlerinin çıkarılması R^2 değerini sırasıyla 0,399 ve 0,315 azaltmış; bu modellerde hasta sayısının pozitif katkısı gözlenmiştir. ANN, hasta sayısı değişkenleri çıkarıldığında kararsız bir performans göstermiştir ($CV R^2=-0,153$). Bununla birlikte bu sonuç, modelin az sayıda özellikle çalışırken yaşadığı optimizasyon güçlüğünden de kaynaklanabilir ve tek başına hasta sayısının “güçlü bir gösterge” olduğu sonucunu desteklemez. Random Forest özelinde negatif artımlı katkı, malzeme adı ile takvim değişkenlerinin ürüne özgü tüketim örüntülerini büyük ölçüde yakaladığını düşündürmektedir. Bu sonuç, permütasyon önemi analiziyle de uyumludur.

Ürün bazlı performans

Toplu R^2 değeri, malzeme adı değişkeninin one-hot kodlamayla modele alınması nedeniyle ürünler arası ölçek farklılıklarından güçlü biçimde etkilenmektedir. Enjektör ve serum seti gibi yüksek hacimli ürünlerin tüketim düzeyi, düşük hacimli ürünlere göre onlarca kat daha büyüktür. Bu nedenle her malzeme için final test döneminde ayrı ayrı hesaplanan R^2 ve WAPE değerleri Tablo 4'te sunulmuştur

(Varyant B, Random Forest modeli). Analiz, aylar arası değişkenliğin ürün düzeyinde ne ölçüde açıklanabildiğini göstermektedir.

Tablo 4: En İyi Modelin (Random Forest, Varyant B) Ürün Bazlı Nihai Test Performansı

Malzeme	Test R ² (ürün bazında)	Test WAPE (%)
İntravenöz Kanül Portlu Mavi No:22	-8,443	56,60
Enjektör 10ml Üç Parçalı Yeşil 21G	-2,454	15,96
Enjektör İnsülin Tek Kullanımlık 26G 100IU 1ML	-1,691	119,14
Serum Seti İğneli	-1,521	19,14
Nebülizer Set Maske Yetişkin	-0,623	57,72
İntravenöz Kanül Portlu Mor No:26	-0,342	78,23
Nebülizer Set Maske Pediatrik	-0,237	95,84
İntravenöz Kanül Portlu Pembe No:20	-0,082	9,35
İntravenöz Kanül Portlu Sarı No:24	0,128	17,22
EKG Elektrot Yetişkin	0,243	13,41

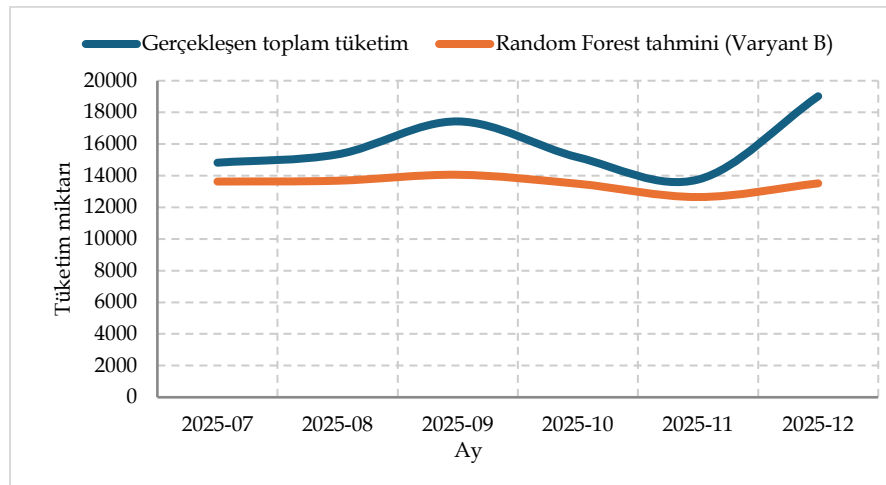
Kaynak: Yazarlar tarafından üretilmiştir.

Tablo 4'te görüldüğü üzere, seçilen Random Forest modelinde 10 malzemenin 8'inde ürün bazlı R² negatiftir. Bu durum modelin, ilgili malzemenin ay bazındaki tüketim dalgalanmasını o malzemenin tarihsel ortalamasından daha iyi tahmin edemediğini göstermektedir. Tablo 3b'de raporlanan toplu test R² değerinin (0,887) büyük ölçüde ürünler arası ölçek farkından kaynaklandığı, hasta başvuru sayısının ürün içi aylık dalgalanmayı açıklama gücünün ise sınırlı kaldığı anlaşılmaktadır. Bu sonuç, Tablo 3d'deki değişken çıkarma bulgusuyla da uyumludur. Bu nedenle hasta başvuru sayısı ile tüketim arasındaki ilişki kesin bir katkı ilişkisi olarak değil, toplu düzeyde gözlenen potansiyel bir ilişki olarak değerlendirilmelidir.

Bu bulgu permütasyon önemi analiziyle de uyumludur. Random Forest modeli üzerinde final test setinde yapılan analizde malzeme adı değişkeninin karıştırılması, ortalama test R² değerinde 1,55 birimlik düşüşe yol açmıştır. Lag-1 hasta sayısı değişkeninin karıştırılması ise yalnızca 0,013 birimlik sınırlı bir düşüş oluşturmuş; üç aylık hareketli ortalama hasta sayısı değişkeninin karıştırılması ölçülebilir bir etki yaratmamıştır (-0,0006). Permütasyon önemi analizinde lag-1 değişkeninin küçük de olsa pozitif katkı göstermesi ile değişken çıkarma analizindeki sonuç arasındaki farklılık, iki yöntemin farklı yapıları ölçmesinden kaynaklanmaktadır. Permütasyon önemi diğer değişkenler modelde sabitken tek bir değişkenin karıştırılmasının etkisini ölçerken, değişken çıkarma analizinde model ilgili değişken olmadan yeniden eğitilmektedir. Her iki analiz de hasta başvuru sayısının modelin açıklama gücüne katkısının malzeme adı değişkenine kıyasla oldukça sınırlı olduğunu göstermektedir.

Nihai test dönemi tahminleri

Final test dönemi olan 2025 Temmuz-Aralık aylarında, en iyi model Random Forest (Varyant B) ile üretilen toplam tüketim tahminleri, gerçekleşen toplam tüketim miktarlarıyla genel eğilim açısından uyumlu bir örüntü göstermiştir (Test R²=0,887). Şekil 1'de final test döneminde ay bazında gerçekleşen ve tahmin edilen toplam tüketim miktarları sunulmaktadır.



Şekil 1: Nihai Test Döneminde Gerçekleşen ve Tahmin Edilen Toplam Tüketim Miktarı

Kaynak: Yazarlar tarafından üretilmiştir.

Şekil 1, modelin toplam tüketim eğilimini genel olarak yakalayabildiğini, ancak Aralık ayındaki keskin artışı olduğu gibi öngöremediğini göstermektedir. Tablo 4'te sunulan ürün bazlı analiz, bu toplu uyumun büyük ölçüde ürünler arası ölçek farkından kaynaklandığını; intravenöz kanül, enjektör veya insülin enjektörü gibi ürünlerde ay bazlı tüketim dalgalanmasının hasta başvuru sayısı ile sınırlı düzeyde açıklanabildiğini ortaya koymaktadır. Bu ürünlerde tüketim, toplam hasta başvuru sayısından çok vaka profiline, klinik uygulama protokollerine ve mevsimsel hastalık yüküne bağlı olabilir. Bu nedenle gelecek çalışmalarda vaka türü, triyaj düzeyi, işlem sayısı ve birim içi protokol değişiklikleri gibi daha ayrıntılı değişkenlerin modele eklenmesi, özellikle düşük hacimli ürünlerdeki tahmin başarısını artırmak açısından önerilebilir.

Tartışma

Bu çalışma, acil servislerde tıbbi malzeme tüketiminin toplu düzeyde belirli ölçüde tahmin edilebildiğini, ancak hasta başvuru yoğunluğunun bu performansa sağladığı artımlı katkının seçilen modelde gösterilemediğini ortaya koymaktadır. Ürün düzeyindeki açıklama gücü ise sınırlı kalmıştır. Bulgular, sağlık kurumlarında operasyonel verilerin yalnızca raporlama amacıyla değil, karar destek ve kaynak planlama amacıyla da kullanılabileceğini; ancak model çıktılarının hangi ürünler için güvenilir olduğunun ayrı biçimde değerlendirilmesi gerektiğini ortaya koymaktadır.

Çalışmada en iyi model seçim süreci, yalnızca çapraz doğrulama sonuçlarına dayanan ve önceden tanımlanan bir kurala (CV R² Ortalaması – CV R² Standart Sapması) göre yapılandırılmıştır. Bu protokolle Random Forest en iyi model olarak seçilmiştir (Seçim Skoru=0,7844). Modelin öne çıkmasının nedeni en yüksek ortalama R² değerine sahip olması değil, katlar arası istikrarının daha yüksek olmasıdır (CV R² Standart Sapması=0,0079). Lineer Regresyon final test setinde daha yüksek R² değerine ulaşmıştır (0,921'e karşı 0,887); ancak modellerden birinin final test sonuçlarına göre tercih edilmesi test setinin bağımsızlığını zayıflatacağından yöntemsel olarak uygun değildir. Bu nedenle seçim kuralı analiz öncesinde belirlenmiş ve model seçimi yalnızca çapraz doğrulama sonuçlarına dayandırılmıştır. Tek bir test döneminde gözlenen performans farkının, modellerin genel üstünlüğü hakkında kesin kanıt oluşturamayabileceği ve örnekleme değişkenliğini yansıtabileceği dikkate alınmalıdır.

Değişken çıkarma analizi (Tablo 3d), hasta başvuru sayısının katkısının model ailesine göre farklılaştığını göstermiştir. Hasta sayısı doğrusal modellerde ölçülebilir bir katkı sağlarken, seçilen model dâhil olmak üzere ağaç tabanlı modellerde bu katkı gözlenmemiştir. Bu farklılık, Montero-Manso ve Hyndman (2021) tarafından ortaya konan yerel ve küresel tahmin modelleri çerçevesiyle ilişkilendirilebilir. Bu çalışmada kullanılan modeller, malzeme adını ortak bir kategorik değişken olarak kullanan küresel modellerdir. Küresel modeller heterojen ürün gruplarında toplu düzeyde güçlü performans gösterebilse de bireysel serilerin zayıf tahmin edilmesini gizleyebilir; Tablo 4'te gözlenen örüntü bu duruma örnektir. Ayrıca düşük hacimli ve aralıklı talep gösteren ürünlerde gözlenen zayıf performans, Liu'nun (2020) kısa yaşam döngülü tıbbi sarf malzemeleri için ele aldığı aralıklı talep tahmin yaklaşımlarıyla ilişkilendirilebilir. Bu ürünlerde Croston ve türevleri gibi aralıklı talebe özgü yöntemlerin test edilmesi gelecek çalışmalar için önerilmektedir.

Varyant A ve Varyant B'nin değişken çıkarma sonuçları, hasta başvuru değişkenlerinin katkısının kullanılan özellik kümesine ve algoritmaya göre değişebildiğini göstermektedir. Yalnızca lag-1 değişkeninin kullanıldığı Varyant A'da hasta sayısının çıkarılması ağaç tabanlı modellerin performansını düşürürken, lag-1 ve roll3 değişkenlerinin birlikte kullanıldığı Varyant B'de aynı katkı gözlenmemiştir. Bu farklılık, önceki üç aya ait hareketli ortalamanın ağaç tabanlı modeller açısından ek bilgi sağlamamış veya sınırlı veri yapısında model karmaşıklığını artırmış olabileceğini düşündürmektedir. Bununla birlikte bu yorum, 48 aylık kısa gözlem dönemi nedeniyle ihtiyatla değerlendirilmelidir.

Ürün bazlı analiz (Tablo 4), çalışmanın dikkat çekici bulgularından birini oluşturmaktadır. On malzemenin sekizinde model, ilgili ürünün tarihsel ortalamasından daha iyi tahmin üretememiştir (negatif R²). Bu durum, toplu R² değerinin büyük ölçüde ürünler arası tüketim hacmi farkından kaynaklandığını ve hasta başvuru sayısının ürün içi aylık dalgalanmayı açıklama gücünün sınırlı kaldığını göstermektedir. Dolayısıyla acil servis başvuru yoğunluğu ile tıbbi malzeme tüketimi arasındaki ilişki, toplu düzeyde gözlenen potansiyel bir ilişki olarak değerlendirilmelidir; hasta başvuru sayısı belirli bir ürünün aylık stok planlaması için tek başına yeterli bir değişken değildir. Bu bulgu, toplu ve alt düzey tahminlerin birlikte değerlendirilmesi gerektiğini vurgulayan çoklu ve hiyerarşik zaman serisi yaklaşımlarıyla uyumludur.

Bulgular yönetsel açıdan, model çıktılarının doğrudan ürün bazlı sipariş veya güvenlik stoku kararı üretmek amacıyla kullanılmaması gerektiğini göstermektedir. Model, toplam tüketim eğilimlerinin

izlenmesi, bütçeleme ve desteklenmesi ve olağandışı sapma gösteren ürünlerin uzman incelemesine yönlendirilmesi amacıyla yardımcı bir karar destek aracı olarak değerlendirilebilir. Tahmin sonuçları, satın alma, depo ve klinik birimler tarafından mevcut stok düzeyi, tedarik süresi ve klinik gereksinimlerle birlikte değerlendirilmelidir. Stok yetersizliği, aşırı stok ve tıbbi atık üzerindeki olası etkiler bu çalışmada doğrudan ölçülmediğinden, bu yararların gelecekteki kurumsal uygulamalarla ayrıca sınanması gerekmektedir.

Çalışmanın bulguları, tıbbi malzeme yönetiminde veri odaklı yaklaşımların maliyet etkinliği ve hizmet sürekliliği açısından önemini vurgulayan literatürle uyumludur (Landry ve Beaulieu, 2013; Saha ve Ray, 2019). Sağlık lojistiği ve tedarik zinciri çalışmalarında doğru tahminin stok maliyetlerini azaltma, hizmet düzeyini artırma ve kaynak kullanımını iyileştirme potansiyeline sahip olduğu belirtilmektedir. Bu çalışma, söz konusu yaklaşımı acil servis özelinde sınavarak literatüre uygulama odaklı bir katkı sunmaktadır.

Sonuç ve öneriler

Bu araştırmada, bir özel hastanenin acil servisinde kullanılan yüksek tüketimli tıbbi sarf malzemelerinin aylık tüketim miktarları makine öğrenmesi ve klasik zaman serisi yöntemleriyle değerlendirilmiştir. Analizler, hasta başvuru sayısının modele katkısının kullanılan algoritmaya göre değiştiğini; seçilen Random Forest modelinde ise ölçülebilir bir artımlı katkı göstermediğini ortaya koymuştur.

Çalışmada karşılaştırılan modeller arasında Random Forest, yalnızca çapraz doğrulama sonuçlarına dayanan ve önceden tanımlanan seçim protokolüyle en iyi model olarak belirlenmiştir. Modelin final test setindeki R^2 değeri 0,887'dir; bu değer model seçiminde kullanılmamış ve yalnızca bağımsız doğrulama amacıyla raporlanmıştır. Bununla birlikte ürün bazlı analiz ve değişken çıkarma analizi, toplu başarının büyük ölçüde ürünler arası tüketim hacmi farkından kaynaklandığını, hasta başvuru sayısının seçilen modelde ölçülebilir artımlı katkı sağlamadığını ve tek bir ürünün aylık tüketimiyle ilişkisinin sınırlı kaldığını göstermiştir. Bu nedenle model, toplam sarf malzeme bütçeleme ve insan denetimli karar desteğinde yardımcı bir araç olarak değerlendirilebilir; ancak ürün bazlı otomatik sipariş kararları için tek başına yeterli değildir. Maliyet, tıbbi atık ve hizmet sürekliliğine ilişkin olası yararlar doğrudan analiz edilmediğinden potansiyel etkiler olarak sunulmaktadır.

Hastane yöneticileri açısından bulgular, model çıktılarının doğrudan sipariş miktarı, güvenlik stoku veya ürün bazlı yenileme kararı üretmek amacıyla kullanılmaması gerektiğini göstermektedir. Model, toplam tüketim eğiliminin izlenmesi, aylık bütçeleme ve desteklenmesi ve olağandışı sapma gösteren ürünlerin uzman incelemesine yönlendirilmesi amacıyla yardımcı bir karar destek aracı olarak değerlendirilebilir. Maliyet, stok tükenmesi, tıbbi atık ve hizmet sürekliliği üzerindeki etkiler bu çalışmada doğrudan ölçülmediğinden, bu alanlardaki olası yararlar ancak gelecekteki kurumsal uygulamalarla sınanabilir.

Çalışmanın bazı sınırlılıkları bulunmaktadır. Veriler tek bir özel hastaneye aittir ve bu nedenle bulguların farklı ölçek ve mülkiyet yapısına sahip hastanelere doğrudan genellenmesi sınırlıdır. Modelde vaka türü, triyaj düzeyi, işlem sayısı, tedarik süresi, stok seviyesi ve birim maliyet gibi değişkenler yer almamaktadır. Ayrıca analizler aylık düzeyde yürütülmüştür; günlük veya haftalık veriler daha ayrıntılı tüketim örüntülerini ortaya koyabilir. Son olarak, model performansı yalnızca 2025 Temmuz-Aralık dönemini kapsayan tek bir 6 aylık final test dönemi üzerinden değerlendirilmiştir; bu durum sonuçların farklı dönemlere genellenebilirliğini sınırlamakta olup, gelecekte kayan pencere (rolling-origin) değerlendirmesiyle daha sağlam biçimde sınanmalıdır. Ayrıca veri seti yalnızca 48 farklı ay ve 10 üründen oluşmakta; çapraz doğrulamada kullanılan üç fold ve tek bir altı aylık final test dönemi, model performansının istatistiksel sağlamlığını sınırlamaktadır. Bu kısıtlar nedeniyle raporlanan performans farkları (özellikle modeller arası küçük farklar) örnekleme gürültüsünden belirgin biçimde ayırtılamamaktadır; gelecek çalışmalarda daha uzun gözlem dönemleri ve çoklu final test pencereleriyle (rolling-origin) tekrarlanan değerlendirme, bulguların sağlamlığını artıracaktır.

Gelecek çalışmalarda triyaj düzeyi, tanı grubu, işlem hacmi, tedarik süresi, stok seviyesi ve mevsimsel hastalık göstergeleri gibi değişkenlerin modele eklenmesi; günlük veya haftalık verilerle, daha uzun dönemlerde ve birden fazla hastanede değerlendirme yapılması önerilmektedir. Bu çalışma, acil servis sarf malzeme tüketiminin toplu düzeyde veri temelli olarak izlenebileceğini, ancak ürün bazlı otomatik kararlar için daha ayrıntılı veri ve doğrulamaya ihtiyaç bulunduğunu göstermektedir.

Bazı uygulama alanlarında basit ve yorumlanabilir modeller, karmaşık modellere göre daha uygun olabilir. Bu çalışmada doğrusal modellerin bazı ölçütlerde rekabetçi performans göstermesi de bu değerlendirmeyi desteklemektedir. Bununla birlikte model seçimi yalnızca çapraz doğrulama

performansı ve katlar arası istikrarı birlikte değerlendiren önceden tanımlanmış Seçim Skoruna dayandırılmış; final test sonuçları model seçiminde kullanılmamıştır.

Hakem Değerlendirmesi / Peer-review:

Dış bağımsız

Externally peer-reviewed

Çıkar Çatışması / Conflict of Interests:

Yazarlar çıkar çatışması bildirmemiştir.

The authors have no conflict of interest to declare.

Finansal Destek / Grant Support:

Yazarlar bu çalışma için finansal destek almadıklarını beyan etmiştir.

The authors declared that this study received no financial support.

Etik Beyan / Ethical Statement:

Araştırmada yalnızca malzeme-ay düzeyinde toplulaştırılmış tüketim ve başvuru verileri kullanılmış; bireysel hasta verisi, kişisel sağlık verisi, anket, görüşme veya klinik müdahale yer almamıştır. Çalışma, ilgili özel hastane yönetiminin 10.01.2025 tarihli ve 2025/112 sayılı araştırma izni kapsamında yürütülmüştür. Bu nedenle etik kurul onayı gerekmemektedir.

Only aggregated material-month-level consumption and visit data were used; no individual patient data, personal health data, surveys, interviews, or clinical interventions were involved. The study was conducted under the research permission dated 10 January 2025 and numbered 2025/112 issued by the relevant private hospital administration. Therefore, ethics committee approval was not required.

Yazar Katkıları / Author Contributions:

Fikir/Kavram/Tasarım: E.Ç.; Veri Toplama ve/veya İşleme: E.Ç.; Analiz ve/veya Yorum: E.Ç., R.S.D.; Kaynak Taraması: E.Ç.; Makalenin Yazımı: E.Ç.; Eleştirel İnceleme: A.M., R.S.D.; Onay: E.Ç., A.M., R.S.D.

Idea/Concept/Design: E.Ç.; Data Collection and/or Processing: E.Ç.; Analysis and/or Interpretation: E.Ç., R.S.D.; Literature Review: E.Ç.; Writing the Article: E.Ç.; Critical Review: A.M., R.S.D.; Approval: E.Ç., A.M., R.S.D.

Kaynakça / References

- Ali, M. K. M., Rabiü, S., & İsmail, M. T. (2025). *Analytical inventory management and optimization: Theories, methods and applications* (1st ed.). CRC Press. <https://doi.org/10.1201/9781003534860>
- Altın, F. G., Budak, İ., & Özcan, F. (2023). Predicting the amount of medical waste using kernel-based SVM and deep learning methods for a private hospital in Turkey. *Sustainable Chemistry and Pharmacy*, 33, Article 101060. <https://doi.org/10.1016/j.scp.2023.101060>
- Asheim, A., Bjørnsen, L. P. B. W., Næss-Pley, L. E., Uleberg, O., Dale, J., & Nilsen, S. M. (2019). Real-time forecasting of emergency department arrivals using prehospital data. *BMC Emergency Medicine*, 19, Article 42. <https://doi.org/10.1186/s12873-019-0256-z>
- Ayer, T., White, C. C., & Zhang, C. (2023). Healthcare inventory management. In J. J. Song (Ed.), *Research handbook on inventory management* (pp. 431–454). Edward Elgar Publishing. <https://doi.org/10.4337/9781800377103.00027>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>

- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer. <https://doi.org/10.1007/978-1-4614-7138-7>
- Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., & Snow, G. L. (2008). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2), 159–170. <https://doi.org/10.1111/j.1553-2712.2007.00032.x>
- Khokhar, S. A. (2023). The challenges of inventory management in medical supply chain. *South Asian Journal of Operations and Logistics*, 2(2), 1–17. <https://doi.org/10.57044/SAJOL.2023.2.2.2306>
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Landry, S., & Beaulieu, M. (2013). The challenges of hospital supply chain management, from central stores to nursing units. In B. T. Denton (Ed.), *Handbook of healthcare operations management: Methods and applications* (pp. 465–482). Springer. https://doi.org/10.1007/978-1-4614-5885-2_18
- Langabeer, J. R., & Helton, J. (2021). *Health care operations management: A systems perspective*. Jones & Bartlett Learning.
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Liu, P. (2020). Intermittent demand forecasting for medical consumables with short life cycle using a dynamic neural network during the COVID-19 epidemic. *Health Informatics Journal*, 26(4), 3106–3122. <https://doi.org/10.1177/1460458220954730>
- Montero-Manso, P., & Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting*, 37(4), 1632–1653. <https://doi.org/10.1016/j.ijforecast.2021.03.004>
- Saha, E., & Ray, P. K. (2019). Modelling and analysis of inventory management systems in healthcare: A review and reflections. *Computers & Industrial Engineering*, 137, Article 106051. <https://doi.org/10.1016/j.cie.2019.106051>
- Sayın, H. C. (2016). Envanter yönetimi. In D. Paşaoğlu (Ed.), *Depolama & envanter yönetimi* (pp. 116–139). Anadolu Üniversitesi Yayınları.
- Tuominen, J., Lomio, F., Oksala, N., Palomäki, A., Peltonen, J., Huttunen, H., & Roine, A. (2022). Forecasting daily emergency department arrivals using high-dimensional multivariate data: A feature selection approach. *BMC Medical Informatics and Decision Making*, 22, Article 134. <https://doi.org/10.1186/s12911-022-01878-7>
- Türkiye İstatistik Kurumu. (2024). *Sağlık harcamaları istatistikleri, 2023* [Haber bülteni]. <https://data.tuik.gov.tr/Bulten/Index?p=Saglik-Harcamalari-Istatistikleri-2023-53561>
- Vollmer, M. A., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., Klaber, R., Cooke, G., Flaxman, S., & Bhatt, S. (2021). A unified machine learning approach to time series forecasting applied to demand at emergency departments. *BMC Emergency Medicine*, 21(1), Article 9. <https://doi.org/10.1186/s12873-020-00395-y>