



ÖZET

Bu çalışmanın temel amacı; Türkiye finansal piyasalarının önemli göstergeleri olan **BİST 100 Endeksi**, **Dolar/TL**, **Euro/TL** ve **Altın/TL** varlıklarının fiyat hareketlerini **Markov Zinciri Analizi** yöntemiyle modellemek ve zaman ölçeğindeki değişimin (1 yıl ve 2 yıl) varlıkların davranışsal dinamikleri üzerindeki etkisini incelemektir. Çalışma kapsamında, Yahoo Finance veri tabanından elde edilen günlük kapanış fiyatları kullanılarak Geçiş Olasılıkları Matrisleri, Uzun Dönem Denge Vektörleri ve Ortalama Kalış Süreleri hesaplanmıştır. Elde edilen bulgular, **BİST 100** endeksinin her iki zaman diliminde de yüksek volatilité ve kısa trend döngüleri (ortalama 2 gün) sergilediğini; buna karşılık **Dolar/TL** kurunun özellikle kısa vadede yükseliş trendini koruma eğiliminin yüksek olduğunu (3.70 gün) ortaya koymuştur. Altın ve Euro ise zamansal değişimlerden bağımsız, kararlı bir yapı sergilemiştir.

GİRİŞ

Finansal piyasalar, makroekonomik gelişmelerden ve küresel risk iştahından etkilenen stokastik bir yapıya sahiptir. Yatırımcılar için fiyatların gelecekteki yönünü öngörmek hayati önem taşır. Bu çalışmada, geleneksel yöntemlerin aksine, bir sistemin gelecekteki durumunun yalnızca mevcut durumuna bağlı olduğu varsayımına dayanan "**Markov Özelliği**" (Markov Property) kullanılmıştır. Bu doğrultuda çalışmanın amacı, Türkiye finansal piyasalarının en önemli göstergeleri olan BİST 100 endeksi, Dolar/TL, Euro/TL ve Altın/TL varlıklarının fiyat hareketlerini **Markov Zincirleri** yöntemiyle modellemek ve zaman periyodunun (1 yıl ve 2 yıl) geçiş dinamikleri üzerindeki etkisini analiz etmektir. Çalışma kapsamında, ele alınan dört farklı varlığın yükseliş ve düşüş trendleri arasındaki geçiş olasılıkları, uzun dönem denge durumları ve ortalama kalış süreleri hesaplanmıştır.

Veri Hakkında

Çalışmada kullanılan veriler, Yahoo Finance veri sağlayıcısı üzerinden Spyder (Python 3.13) programlama dili ve ilgili finansal veri kütüphaneleri kullanılarak çekilmiştir. Python ile çekilen veriler MS Excel ortamına aktarılarak her bir veri seti için analizler yapılmıştır. Analizlerde varlıkların gün içi dalgalanmaları yerine, trendin yönünü belirlemede daha kararlı bir gösterge olan günlük kapanış fiyatları esas alınmıştır. Toplamda 4 farklı varlık için; 02.12.2024 ile 28.11.2025 tarihleri arasını kapsayan 1 yıllık ve 02.12.2023 ile 28.11.2025 tarihleri arasını kapsayan 2 yıllık olmak üzere, iki farklı zaman dilimine ait 8 ayrı veri seti üzerinde çalışılmıştır.



Varlıklar
BIST 100 endeksi
Dolar/TL
EURO/TL
ALTIN/TL



Zaman Aralığı

1 Yıllık: 2024 - 2025
2 Yıllık: 2023 - 2025



Veri Kaynağı

Yahoo Finance
(Günlük kapanış fiyatları)

BULGULAR

Analizin ilk aşamasında, her bir yatırım aracı için “Yükseliş (Y)” ve “Düşüş (D)” durumları arasındaki günlük geçiş olasılıkları (P) hesaplanmıştır. Aşağıdaki tablolarda 2 yıllık ve 1 yıllık geçiş olasılıkları karşılaştırmalı olarak verilmiştir.

GEÇİŞ OLASILIKLARI (P)

Varlıkların 'Yükseliş' ve 'Düşüş' durumları arasındaki günlük hareketliliği, Geçiş Olasılıkları Matrisi ile modellenmiş ve zamansal değişim (1 Yıl vs 2 Yıl) aşağıda karşılaştırılmıştır.

Tablo: Yatırım Araçlarının 1 ve 2 Yıllık Geçiş Olasılıklarının Karşılaştırılması

Yatırım Aracı	Dönem	P _{YY} (Yükseliş → Yükseliş)	P _{YD} (Yükseliş → Düşüş)	P _{DY} (Düşüş → Yükseliş)	P _{DD} (Düşüş → Düşüş)
BİST100	2 Yıl	0.53	0.47	0.49	0.51
	1 Yıl	0.51	0.49	0.50	0.50
Altın/TL	2 Yıl	0.57	0.43	0.64	0.36
	1 Yıl	0.56	0.44	0.65	0.35
Dolar/TL	2 Yıl	0.68	0.32	0.62	0.38
	1 Yıl	0.73	0.27	0.68	0.32
Euro/TL	2 Yıl	0.55	0.45	0.60	0.40
	1 Yıl	0.54	0.46	0.61	0.39

- Dolar:** Dolar/TL paritesinin 1 yıllık periyotta yükselişini sürdürme olasılığı (P_{YY}) **0.73** ile en yüksek seviyede, 2 yıllık periyotta bu oran **0.68**'e gerilemiştir. Bu durum, kurun kısa vadeli şoklarda yukarı yönlü hareketini koruma eğiliminin (yapışkanlık) daha güçlü olduğunu gösterir.
- BİST 100:** Hisse senedi piyasasında yükseliş ve düşüş olasılıkları her iki dönemde de **0.50** bandında yoğunlaşmıştır. Bu istatistik, BİST 100'ün bir sonraki günkü hareketinin tahmin edilmesinde geçmiş verinin etkisinin zayıf olduğunu ve piyasanın **rassal** bir yapı sergilediğini kanıtlar.
- Altın ve Euro :** Altın/TL ve Euro/TL varlıkları, zaman ölçeği değişse bile geçiş olasılıklarında anlamlı bir sapma göstermeyerek **kararlı (durağan)** bir yapı sergilemiştir.

DENGE DURUM ANALİZİ (π)

Her bir yatırım aracının, başlangıç durumundan bağımsız olarak uzun vadede Yükseliş veya Düşüş durumunda bulunma olasılıkları (π vektörü) hesaplanmıştır. Bu analiz, piyasanın genel eğilimini ortaya koyar.

Tablo: Yatırım Araçlarının Uzun Dönem Denge Vektörlerinin Karşılaştırılması

Yatırım Aracı	2 Yıllık Ortalamada Yükseliş (π_Y)	1 Yıllık Ortalamada Yükseliş (π_Y)
Dolar/TL	%66.0	%71.6
Altın/TL	%59.8	%59.6
Euro/TL	%57.1	%57.0
BİST100	%51.0	%50.5

- Rassal Yürüyüş Hipotezi:** BİST 100 endeksinin uzun vadede yükseliş durumunda bulunma olasılığı **%50.5** (1 yıl) ve **%51.0** (2 yıl) olarak hesaplanmıştır. Bu oranların %50'ye çok yakınsaması, piyasanın uzun vadede belirgin bir yön tayin edemediğini gösterir.
- En Güçlü Yükseliş Eğilimi:** Dolar/TL, **%71.6** (1 yıl) ve **%66.0** (2 yıl) ile uzun vadede yükseliş durumunda kalma olasılığı en yüksek yatırım aracıdır.
- Güvenli Liman Etkisi:** Altın/TL, veri setinin uzunluğundan bağımsız olarak yaklaşık **%60** seviyesindeki denge olasılığını korumuştur. Bu bulgu, altının spekülâtif dalgalanmalardan en az etkilenen, istatistiksel olarak en güvenilir varlık olduğunu doğrular.

ZAMANSAL KARŞILAŞTIRMA (Ψ)

Piyasalardaki trendlerin ne kadar süreyle korunduğunu ölçmek amacıyla Ortalama İlk Geçiş Zamanı Matrisleri (Ψ) hesaplanmıştır. Özellikle “Yükselişten Düşüşe Geçiş Süresi” (τ_{12}), bir boğa trendinin ortalama kaç gün sürdüğünü ifade eder

Tablo: Yatırım Araçlarının Ortalama İlk Geçiş Sürelerinin Karşılaştırılması

Yatırım Aracı	2 Yıllık Veri	1 Yıllık Veri
Dolar/TL	3.13 Gün	3.70 Gün
Altın/TL	2.33 Gün	2.27 Gün
Euro/TL	2.22 Gün	2.17 Gün
BİST100	2.13 Gün	2.04 Gün

- Yüksek Volatilité (BİST 100):** Endeksin bir yükseliş trendini bozup düşüşe geçmesi ortalama **2.04 gün** sürmektedir. Bu sürenin kısalığı, piyasa hafızasının zayıf olduğunu ve trend dönüşümlerinin çok sık yaşandığını gösterir
- Zamansal Duyarlılık (Dolar/TL):** Dolar/TL kurunda yükseliş trendinin ömrü 1 yıllık veride **3.70 gün** iken, 2 yıllık veride **3.13 güne** düşmüştür. Bu durum, döviz kurunun kısa vadeli konjonktürde daha dirençli olduğunu, vade uzadıkça trend momentumunun azaldığını ortaya koymaktadır.
- Orta Vadeli Denge:** Euro ve Altın varlıklarında trend değişim süreleri (sırasıyla **2.22** ve **2.33** gün) zaman ölçeğine göre değişmeyerek piyasa şoklarına karşı daha stabil bir davranış sergilemiştir.

SONUÇ

- BİST 100'de Yüksek Volatilité:** Endeksin trend değişim süresinin çok kısa (ortalama 2 gün) olması, piyasanın "hafızasız" olduğunu ve fiyat hareketlerinin **rassal yürüyüşe** yakınsadığını göstermiştir.
- Dolar/TL'de Yapışkanlık Etkisi:** Dolar kuru, özellikle kısa vadede (1 yıl) yukarı yönlü şoklara karşı dirençli ve "yapışkan" bir trend izlerken, vade uzadıkça bu momentumun azaldığı tespit edilmiştir.
- Altın "Güvenli Liman":** Altın/TL, 1 ve 2 yıllık zaman dilimlerinde en kararlı geçiş olasılıklarını sergileyerek, spekülâtif dalgalanmalardan bağımsız, istikrarlı bir **güvenli liman** olduğunu istatistiksel olarak kanıtlamıştır.
- Yöntem Başarısı:** Analizler, finansal zaman serilerinin **Markov Özelliği** taşıdığını; özellikle döviz ve emtia piyasalarındaki trend devamlılığının bu yöntemle başarılı şekilde modellenebildiğini doğrulamıştır.

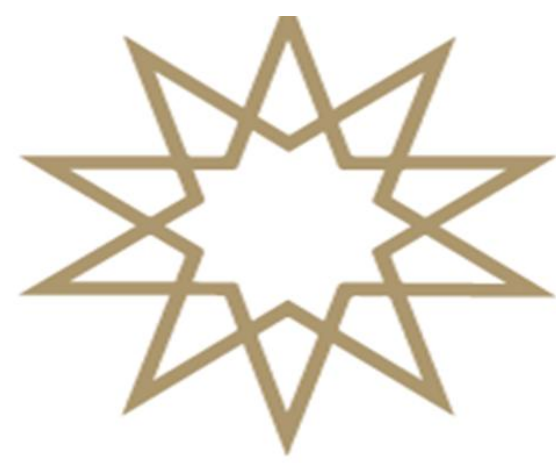


ÖNERİLER

- Yatırımcı Stratejisi (Portföy Çeşitlendirmesi):** BİST 100'ün yüksek volatilitesi karşısında, zaman ölçeğinden bağımsız bir istikrar sunan **Altın ve Euro**, portföylerde dengeleyici ve defansif bir unsur olarak kullanılmalıdır.
- Politika Yapıcılar İçin:** Dolar/TL kurundaki güçlü yukarı yönlü yapışkanlık, kur şoklarının kalıcı etkiler bırakma potansiyelini artırmaktadır. Bu nedenle, volatilitéyi düşürücü para politikası müdahalelerinde **zamanlama** kritik önem taşır.
- Gelecek Çalışmalar (Hibrit Modeller):** BİST 100 gibi rassal hareket eden piyasaların tahmini için, Markov Zincirleri'nin **Yapay Sinir Ağları** ile desteklendiği hibrit modeller veya makroekonomik değişkenlerin (CDS, VIX) dahil edildiği **MS-VAR** modelleri kullanılabilir.

KAYNAKÇA

- [1] Durgun, F., (2023). *Markov Rejim Değişim Modelleri Teori ve Uygulama*.
[2] Şenol, Z., (2020). "Borsa Oynaklığının Markov Rejim Dönüşüm Yöntemiyle Analizi".
[3] Yılmaz, T. & Sevüktekin, M., (2025). *ABD Tahvil Getirisi ve Altın İlişkisinin Analizi*.



Yıldırım

FEN EDEBİYAT FAKÜLTESİ

2025-2026 Yarıyılı İstatistik BÖLÜMÜ

BITCOIN RAINBOW CHART (LOGARİTMİK REGRESYON) YAKLAŞIMININ İSTATİSTİKSEL GEÇERLİLİĞİ VE ZAMAN SERİSİ MODELLERİYLE KARŞILAŞTIRILMASI

Ayşenur ALAHABALI ÖRÜNG 21023607

Doç. Dr. Gülder KEMALBAY

ÖZET

Bu çalışmada Bitcoin kapanış fiyatları kullanılarak (2010–2024) uzun dönem trend yaklaşımı (Rainbow Chart/logaritmik regresyon) ile kısa dönem tahmin modelleri (ARIMA ve Prophet) karşılaştırılmıştır. Durağanlık analizi için ADF testi uygulanmış; ham fiyat serisinin durağan olmadığı, log-getiri serisinin ise güçlü biçimde durağan olduğu bulunmuştur. Tahmin performansı son 365 günlük test penceresinde RMSE/MAE gibi hata ölçütleriyle değerlendirilmiştir. Sonuçlar, uzun dönem bant yapısının fiyat rejimlerini açıklamada yararlı olduğunu; kısa dönem tahminlerde ise model seçiminin veri dönemi ve dönüşümlere duyarlı olduğunu göstermektedir.

AMAÇ

Bitcoin fiyatını hem uzun dönem trend hem de kısa dönem tahmin perspektifiyle incelemek Rainbow, ARIMA, Prophet modellerini aynı veri üzerinde karşılaştırmak.

MATERYAL VE METOT

Veri:

- Günlük Bitcoin kapanış fiyatı (*Close*), 2010–2024
- Olgunluk dönemi analizi: 2014–2024 (tahmin modelleri için)

Ön İşleme:

- Log dönüşüm: $\ln(P_t)$
- Log-getiri: $\Delta \ln(P_t) = \ln(P_t) - \ln(P_{t-1})$
- ADF durağanlık testi

	Seri	ADF	p-value	lag	nobs
0	Close (P_t)	-0.357925	9.168903e-01	33	5025
1	Log_Price ($\ln(P_t)$)	-2.965661	3.822805e-02	23	5035
2	Log_Return ($\Delta \ln(P_t)$)	-12.228044	1.069274e-22	22	5035

Modeller:

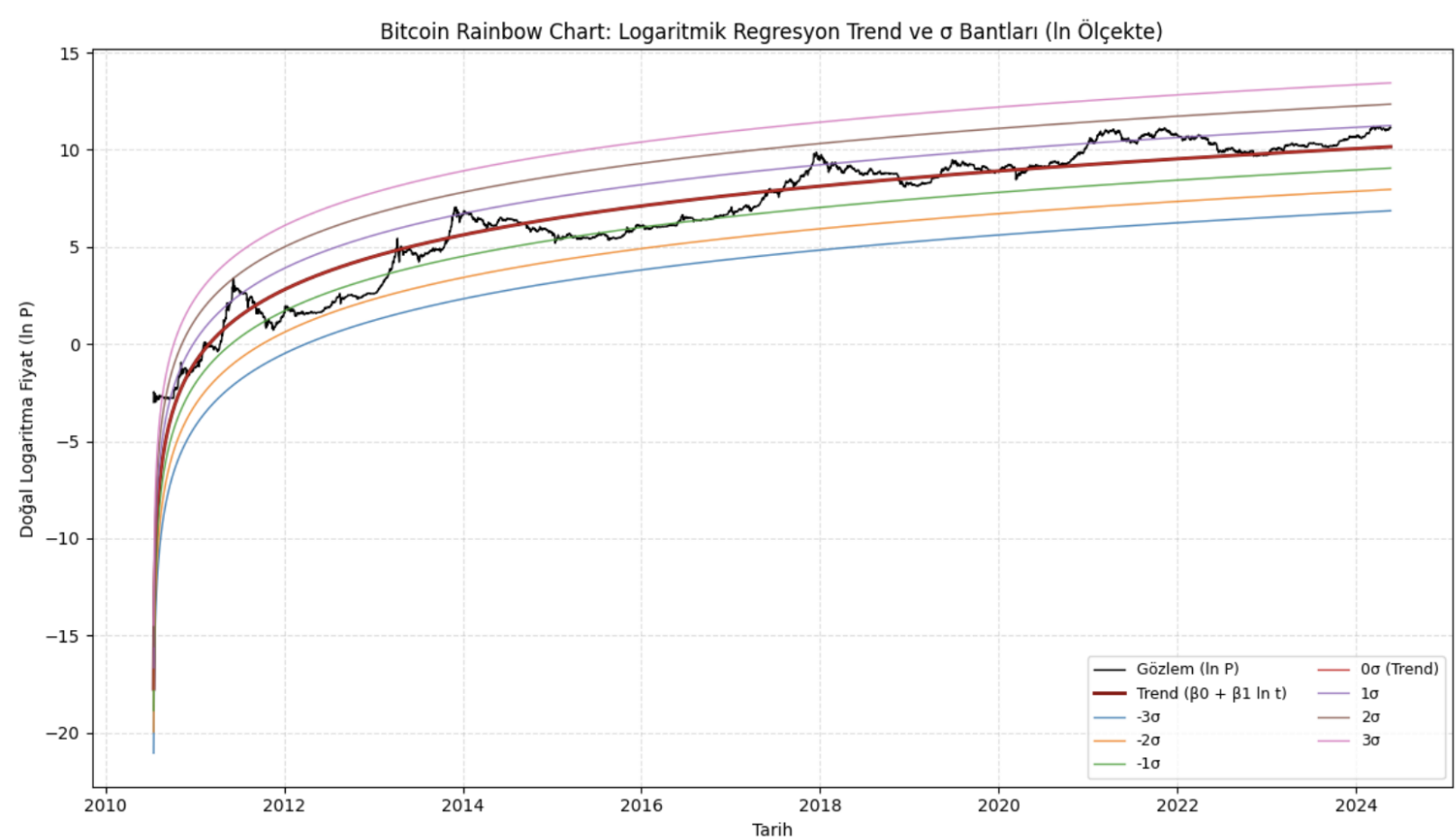
ARIMA(0,1,0)+Drift, Optimize Prophet ve Logaritmik Regresyon.

Testler:

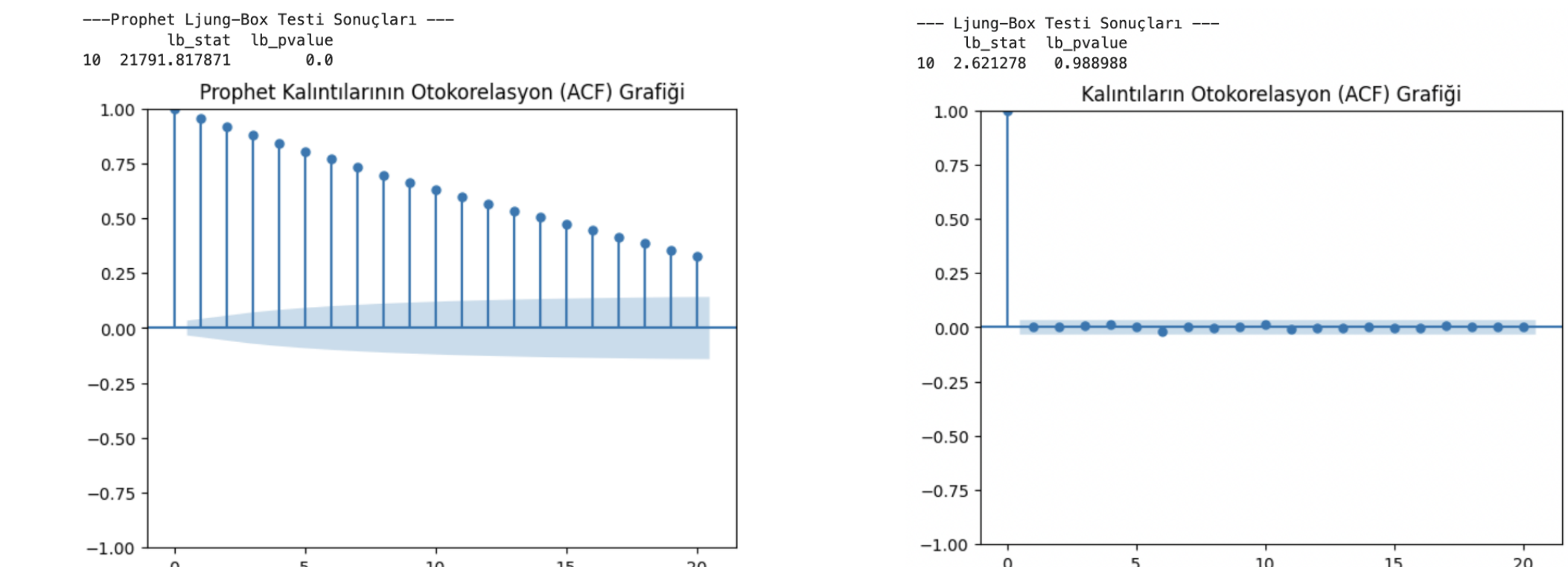
ADF (birim kök), Ljung–Box (otokorelasyon), Diebold–Mariano (tahmin karşılaştırma)

ADF testi sonuçlarına göre ham kapanış fiyatı serisi için birim kök hipotezi reddedilememiş ve seri durağan bulunmamıştır. Buna karşılık log-getiri serisi için birim kök hipotezi güçlü biçimde reddedilmiş olup seri durağan kabul edilmiştir.

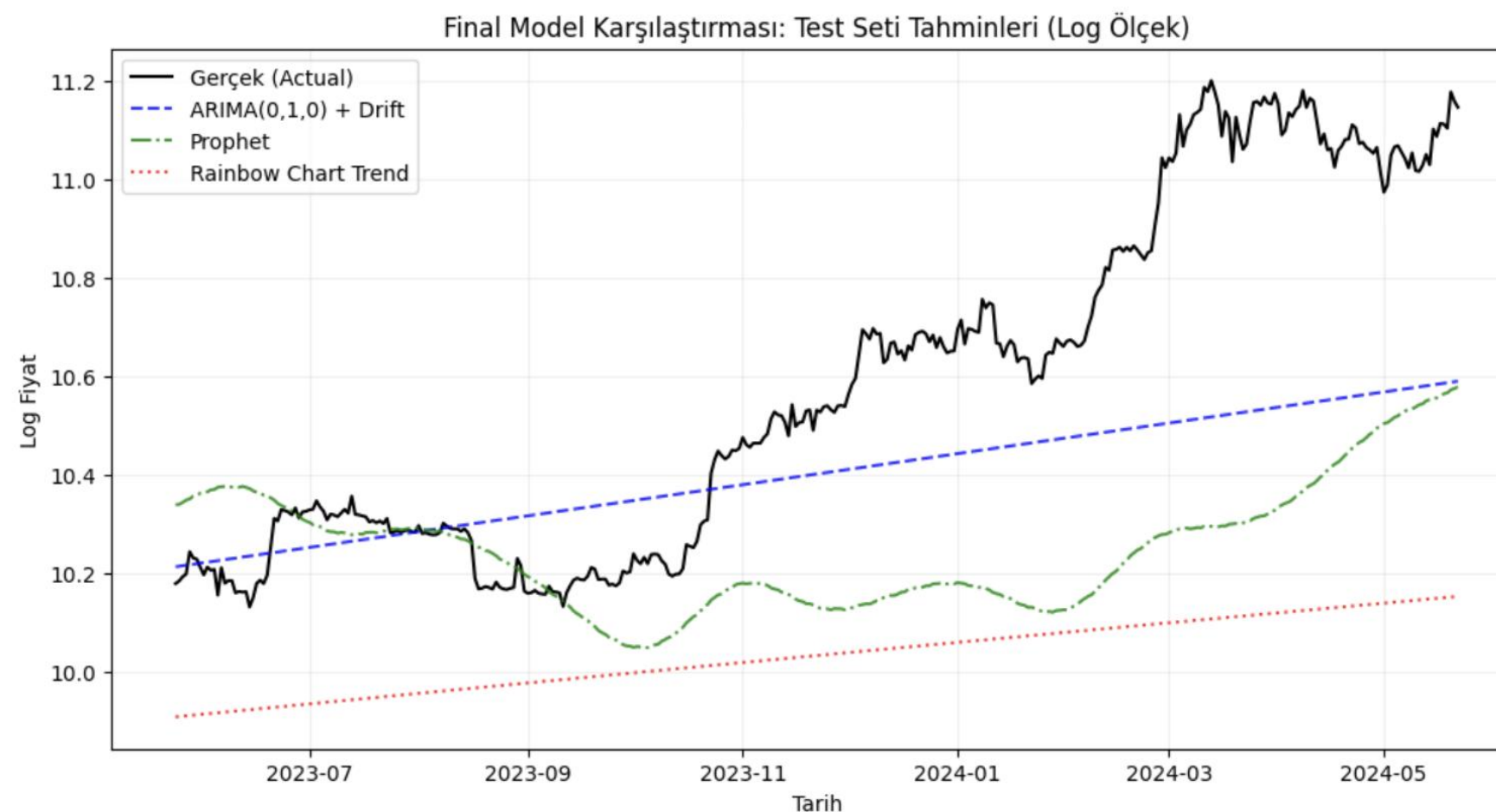
BULGULAR VE GÖRSEL ANALİZ



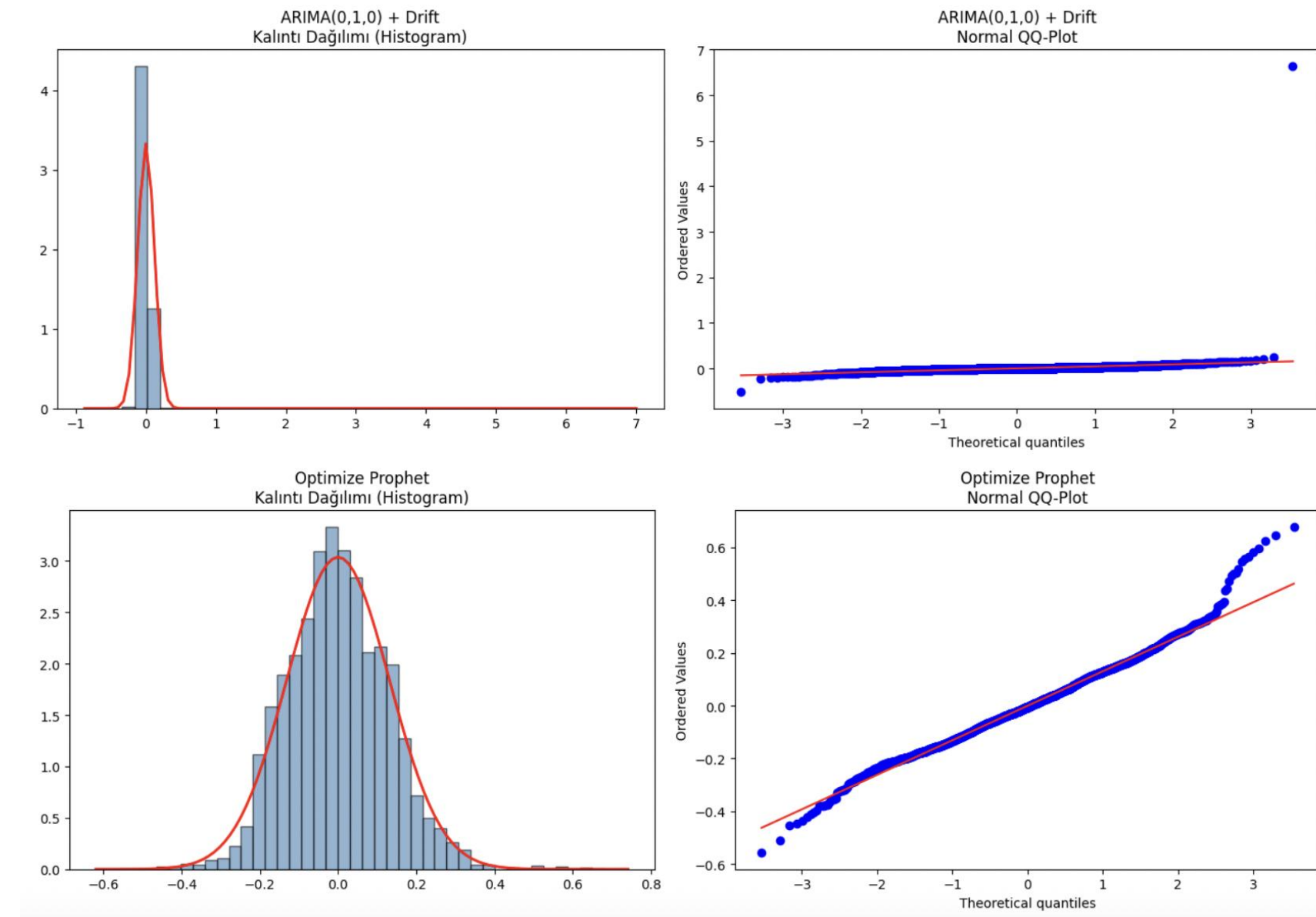
Bitcoin Rainbow Chart (Logaritmik Regresyon) Modeli ve Kalıntı Analizi (2010–2024). Regresyon modeli uzun vadeli büyüme eğilimini (deterministik trend) yüksek R^2 ile açıkla da, kalıntıların geniş bantlarda dalgalanması kısa vadeli tahmin için zaman serisi modellerine ihtiyaç duyulduğunu işaret etmektedir.



Prophet (Sol) ve ARIMA (Sağ) Modelleri Kalıntılarının ACF Karşılaştırması. Prophet kalıntılarındaki güçlü otokorelasyon ($p \approx 0.00$), modelin yapısal bilgiyi tam açıklayamadığını; ARIMA kalıntılarının ise güven sınırları içinde kalarak "beyaz gürültü" şartını sağladığını ($p \approx 0.988$) desteklemektedir.



Test Seti (Son 365 Gün) Tahmin Performansı Karşılaştırması. ARIMA(0,1,0) + Drift modelinin (mavi kesikli çizgi), Bitcoin'in stokastik momentumuna en yakın seyri izleyerek en düşük RMSE (0.304) değerine ulaştığı gözlemlenmektedir.



Modellerin Kalıntı Teşhis Grafiklerinde ARIMA modeline ait hataların normal dağılıma (QQ-Plot referans çizgisi) tam uyum sağladığı, Prophet modelinde ise özellikle uç değerlerde (kuyruklarda) sapmaların olduğu görülmektedir.

SONUÇ

Bu çalışma kapsamında Bitcoin fiyat dinamikleri üzerine yapılan karşılaştırmalı analizler sonucunda şu temel bulgulara ulaşılmıştır:

En Başarılı Model: ARIMA(0,1,0) + Drift modelinin, 0.3044 RMSE değeri ile Bitcoin fiyat öngörüsünde hem Prophet hem de Rainbow Chart modellerine göre istatistiksel olarak anlamlı derecede ($p < 0.001$) daha başarılı olduğu Diebold-Mariano testi ile kanıtlanmıştır.

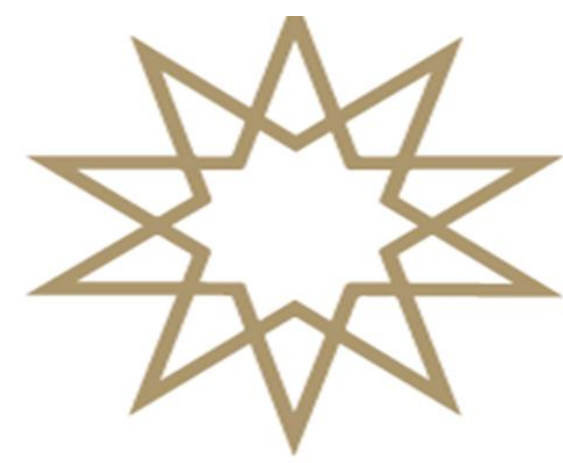
Model Geçerliliği: Teşhis kontrolleri sonucunda, ARIMA modelinin kalıntılarının "beyaz gürültü" (*white noise*) şartını sağladığı; buna karşın daha karmaşık bir yapıya sahip olan Prophet modelinin kalıntılarında güçlü otokorelasyon ($p \approx 0.00$) tespit edildiği saptanmıştır.

Parsimony (Yalınlık) İlkesi: Elde edilen bulgular, Bitcoin gibi yüksek volatiliteye sahip finansal varlıkların modellenmesinde, stokastik momentumu (*drift*) temel alan yalın modellerin, bileşenli ve karmaşık modellere göre daha dirençli (*robust*) tahminler sunduğunu desteklemektedir.

Piyasa Etkinliği: En iyi performansın "Sürüklenmeli Rastgele Yürüyüş" yapısı tarafından sergilenmesi, Bitcoin piyasasının zayıf formda etkin piyasa hipoteziyle uyumlu bir karakteristik sergilediğine işaret etmektedir.

KAYNAKÇA

- P. Choksi, "Bitcoin Historical Dataset (2010-Now)," *Kaggle*
- Nakamoto, S., (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*.
- Hua, Y. (2020). Bitcoin Price Prediction Using ARIMA and LSTM. *E3S Web of Conferences*, 214, 01050



2025-2026 Güz Yarıyılı İSTATİSTİK BÖLÜMÜ

Ekonomik Kriz ve Tüketim: Türkiye’de Ruj Etkisi’nin Zaman Serisi Analizi

Ece YILDIRIM 20023031

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

ÖZET

Bu çalışmanın amacı, 2010-2025 yılları arasında Türkiye ekonomisinde gözlemlenen yüksek enflasyon ve makroekonomik belirsizliklerin, tüketici harcamaları üzerindeki etkisini "Ruj Etkisi" (Lipstick Effect) teorisi çerçevesinde incelemektir. Klasik iktisat teorisi kriz dönemlerinde lüks tüketimin azalmasını öngörürken, bu çalışma Türkiye’de kozmetik harcamalarının reel olarak arttığını savunmaktadır. Çalışmada, TÜİK Perakende Satış Hacim Endeksi ve TCMB verileri kullanılarak Zaman Serisi Analizi ve Ekonometrik Modelleme yöntemleri uygulanmıştır. Bulgular, enflasyon ile kozmetik talebi arasında çok güçlü pozitif bir ilişki ($r=0.93$) olduğunu ve tüketicilerin giyim harcamalarından kıyarak kozmetik ürünlerine yöneldiğini (İkame Etkisi) ortaya koymaktadır. Bu durum, kozmetik tüketiminin Türkiye’de ekonomik bir tercihten öte, borçlanma ile finanse edilen psikolojik bir savunma mekanizması olduğunu göstermektedir.

Ruj Etkisi’nin Zaman Serisi Analizi

Gelişmekte olan ekonomilerde kronik hale gelen enflasyon ve satın alma gücü erozyonu, hanehalkı bütçelerini tehdit etmektedir. Problemin temelini; makroekonomik göstergelerdeki bozulmaya rağmen, kozmetik harcamalarındaki artışın rasyonel bir tercih mi yoksa krizin yarattığı bir "telafi edici tüketim" mi olduğu sorusu oluşturmaktadır. Bu çalışma, söz konusu fenomeni anket verileriyle değil, resmi makro verilerle test ederek literatürdeki metodolojik boşluğu doldurmayı hedeflemektedir.

Ruj Etkisi (The Lipstick Effect): Lauder (2001) tarafından formüle edilen bu kavram, tüketicilerin ekonomik güvensizlik dönemlerinde pahalı lükslerden (otomobil, konut) vazgeçerken, kendilerini iyi hissettirecek "küçük lükslere" yönelmesidir.

İkame Etkisi (Substitution Effect): MacDonald ve Dildar (2020), kozmetik harcamalarındaki artışın, aslında giyim gibi daha pahalı kategorilerden yapılan tasarrufun bir sonucu olduğunu savunur.

Analiz, Türkiye ekonomisindeki yapısal kırılmaları kapsayan 2010-2025 dönemine ait aylık verilerle gerçekleştirilmiştir.

Bağımlı Değişken: TÜİK Kozmetik ve Kişisel Bakım Perakende Satış Hacim Endeksi. (Enflasyondan arındırılmış reel talep).

Bağımsız Değişkenler: TÜFE (Enflasyon), Tüketici Güven Endeksi, Reel Efektif Döviz Kuru

Verilerin analizi dört aşamada gerçekleştirilmiştir:

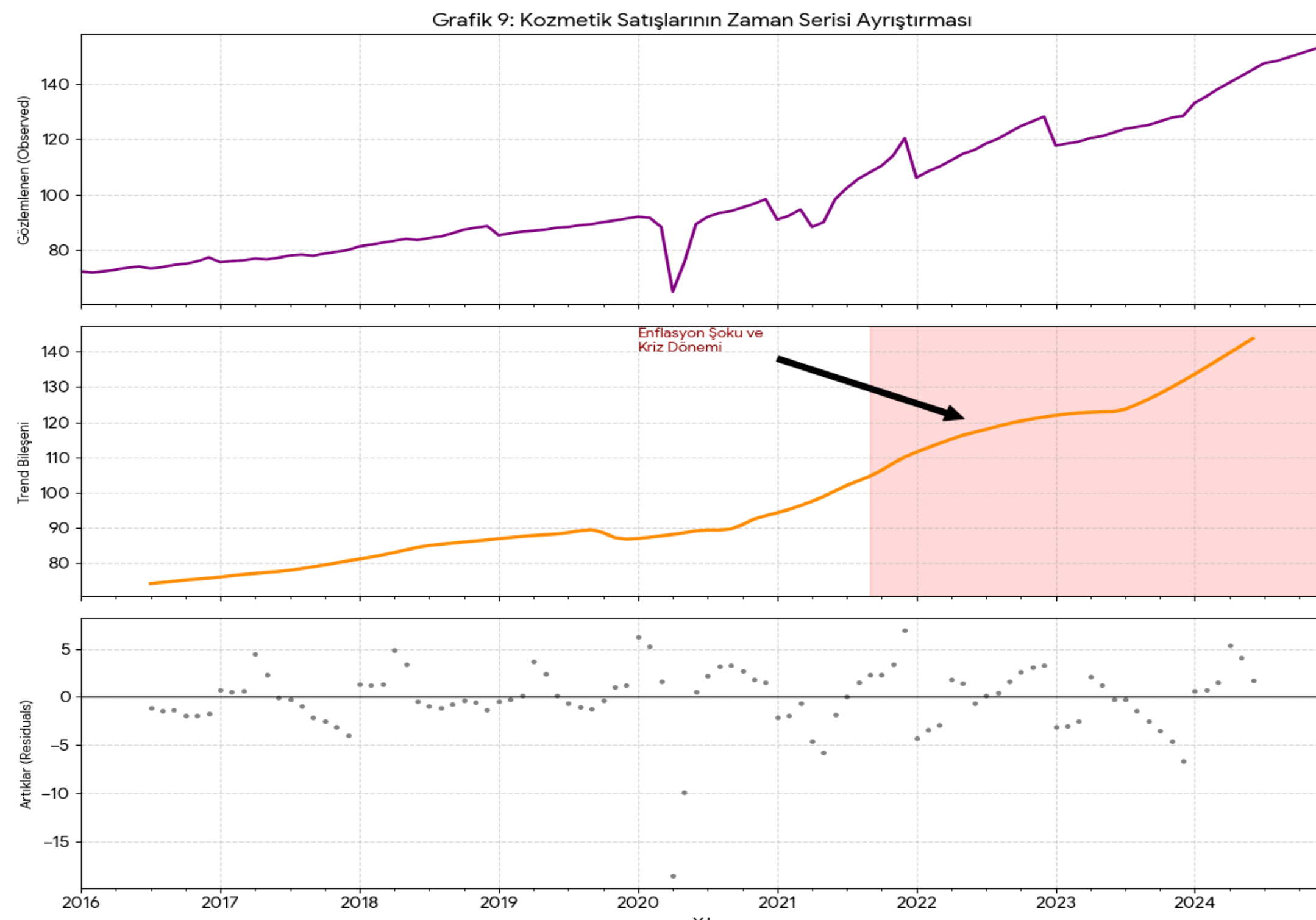
Toplamsal Ayrıştırma (Additive Decomposition): Serinin Trend, Mevsimsellik ve Rassal bileşenlerine ayrılması.

Trend Gücü Analizi: Wang, Smith ve Hyndman (2006) yöntemine göre Trend Gücü katsayısı 0.9889 olarak hesaplanmış, serideki değişimin ana kaynağının uzun vadeli trend olduğu görülmüştür.

Logaritmik Regresyon Modeli: Fiyat esnekliğini ölçmek için kurulan model şu şekildedir;

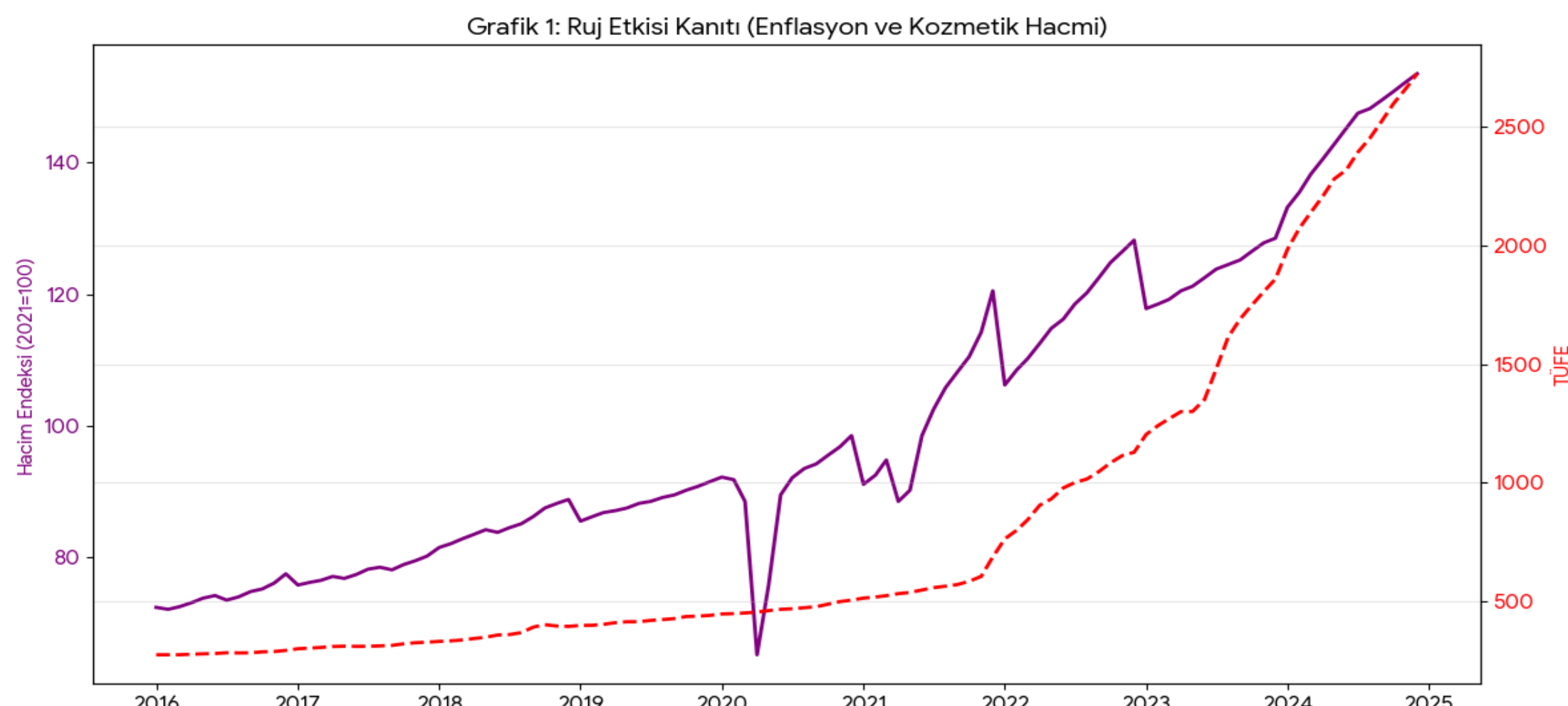
$$\ln(Kozmetik_t) = 2.37 + 0.34 \cdot \ln(TÜFE_t) + \epsilon_t$$

Bu modelin açıklayıcılık katsayısı 0.956 olarak bulunmuştur.



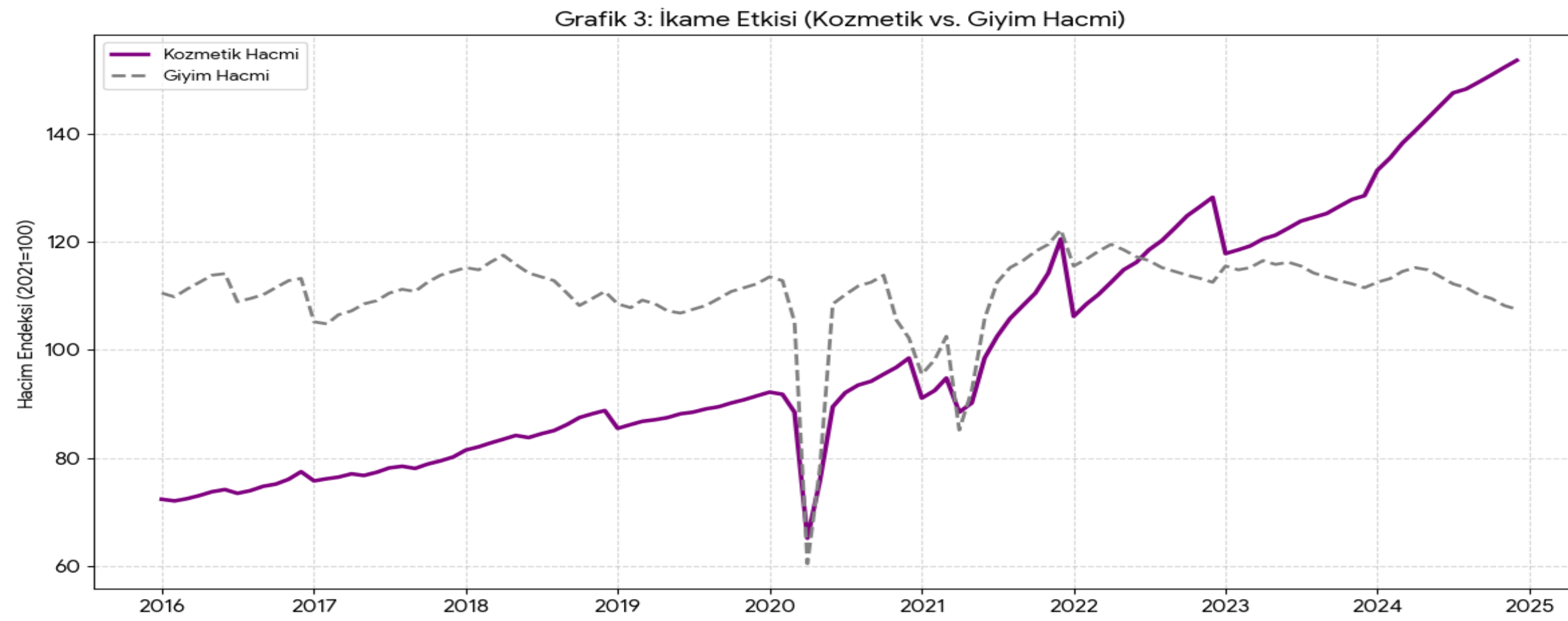
Grafik 9: Zaman Serisi Ayrıştırması ve Yapısal Kırılma

Kozmetik Satış Hacim Endeksi, toplamsal ayrıştırma modeline (Additive Decomposition) tabi tutulmuştur. Wang-Smith-Hyndman yöntemine göre Trend Gücü 0.98 olarak hesaplanmış, serideki değişimin ana kaynağının uzun vadeli trend olduğu görülmüştür. 2021 yılı enflasyon şoku sonrasında trend bileşeninde yukarı yönlü keskin bir Yapısal Kırılma mevcuttur.



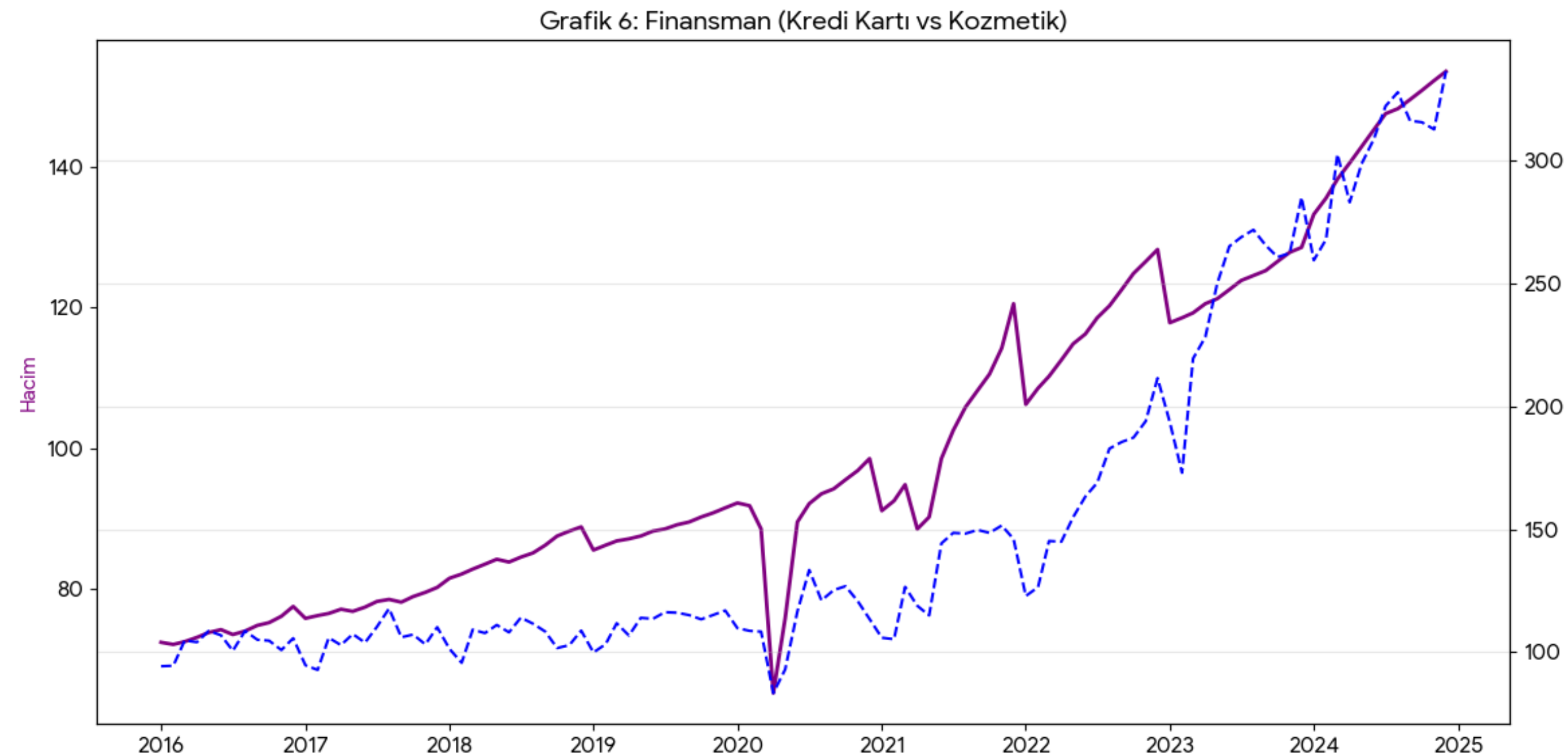
Grafik 1: Ruj Etkisi Paradoksu (Enflasyon vs. Satış Hacmi)

Ekonomik teorinin aksine, Genel Enflasyon (TÜFE) ile Reel Kozmetik Satış Hacmi arasında $r=0.93$ düzeyinde çok güçlü ve pozitif bir korelasyon saptanmıştır $p<0.01$. Fiyatlar genel seviyesindeki artış, talebi düşürmemiş; aksine kozmetik tüketimi reel olarak %64.8 artış göstermiştir.



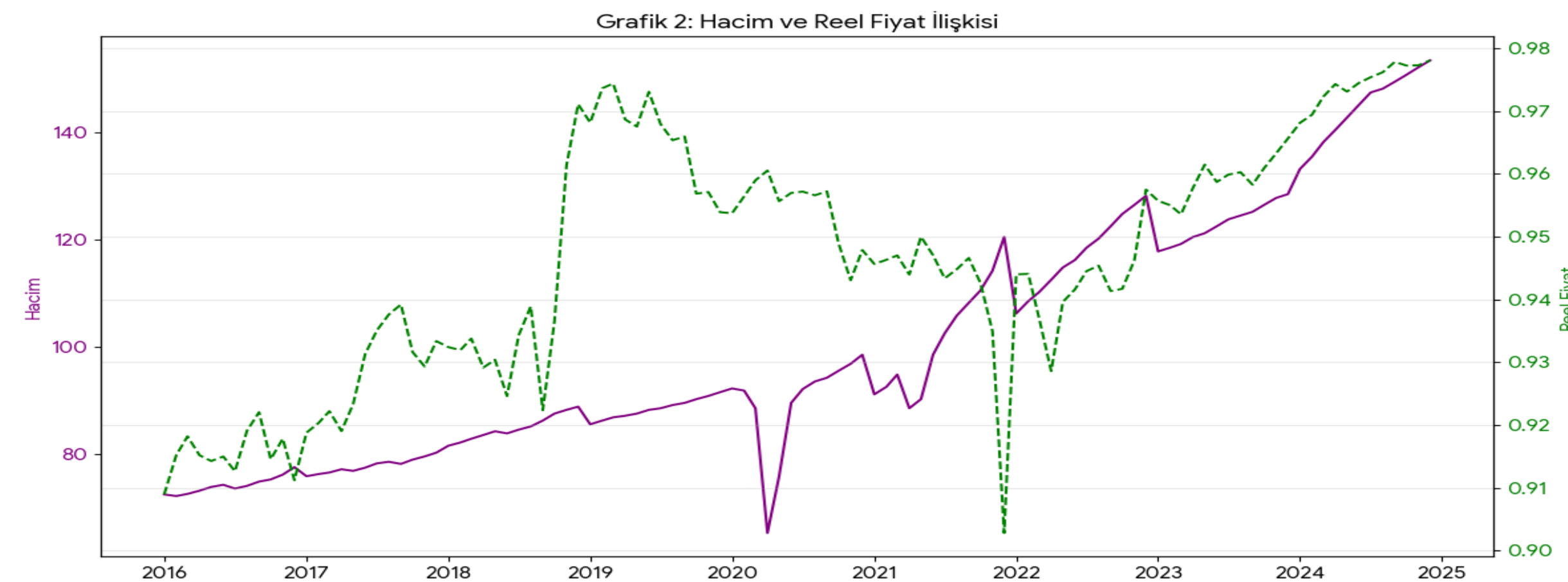
Grafik 3: Tüketicilerin İkame Etkisi (Kozmetik vs. Giyim)

Harcama sepeti analiz edildiğinde; kriz dönemlerinde 'Giyim ve Ayakkabı' harcamaları (Gri Çizgi) reel olarak yatay seyrederken veya küçülürken, 'Kozmetik' harcamaları (Mor Çizgi) dikey bir büyüme trendine girmiştir. Tüketici, yüksek fiyatlı tekstil ürünlerinden kısıtıldığı bütçeyi 'ulaşılabilir lüks' olan kozmetiğe aktarmıştır (Substitution Effect).



Grafik 6: Tüketicinin Finansmanı ve Borçlanma

Kozmetik satış hacmindeki artış, 'Reel Kredi Kartı Harcamaları' ile birebir senkronize hareket etmektedir. Bu bulgu, tüketicilerin alım gücü düşmesine rağmen yaşam standartlarını korumak ve psikolojik tatmin sağlamak amacıyla gelecek gelirlerinden borçlandığını göstermektedir.

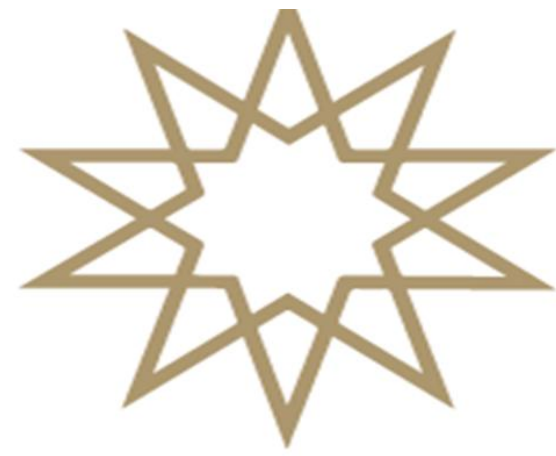


Grafik 2: Fiyat Esnekliği ve Talep Direnci

Kozmetik ürünlerinin reel fiyatı (göreceli pahalılığı) artmasına rağmen, talep düşmemiştir. Bu durum, kozmetik ürünlerinin kriz dönemlerinde 'düşük fiyat esnekliğine' sahip olduğunu ve tüketicinin psikolojik tatmin uğruna fiyat artışlarını tolere ettiğini kanıtlamaktadır.

KAYNAKÇA

- Hill, S. E., et al. (2012). Boosting beauty in an economic decline. JPSP.
- MacDonald, D., & Dildar, Y. (2020). Evidence for the lipstick effect. JBEE.
- TÜİK & TCMB. (2025). Perakende Satış Endeksleri ve EVDS Veri Tabanı.
- Netchaeva, E., & Rees, M. W. (2016). Strategically stunning: The professional motivations behind the lipstick effect. Psychological Science



2025-2026 Güz Yarıyılı İstatistik Bölümü

TÜRKİYE’DE İÇ GÖÇÜN BELİRLEYİCİLERİ: PANEL VERİ ANALİZİ VE YERÇEKİMİ MODELİ

Betül Feyza ÇOKER - 21023043

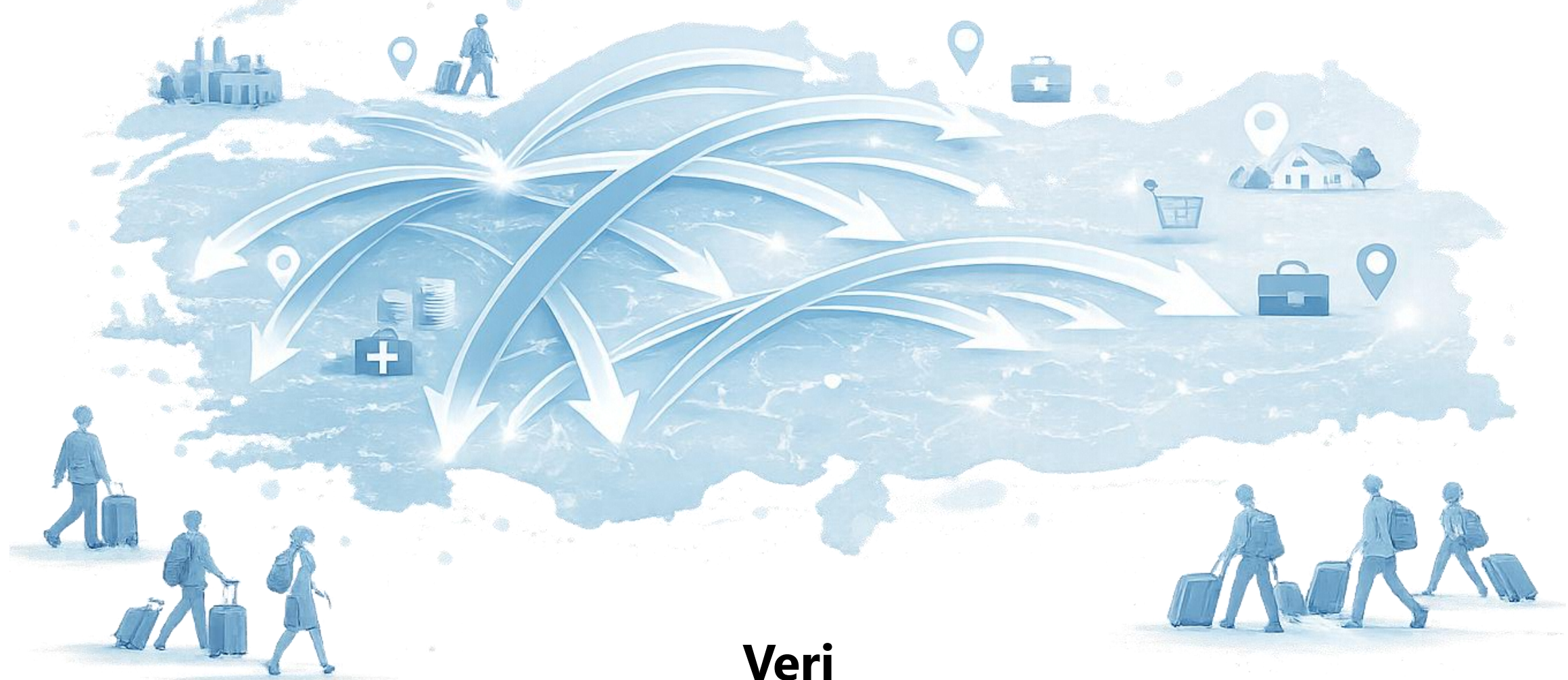
Danışman: Doç. Dr. Erhan ÇENE

ÖZET

Günümüzde insanların göç etme nedenleri farklılık göstermektedir. Bu nedenlere bağlı olarak göç edilen yerler de değişmektedir. İnsanların bulundukları ve göç edecekleri bölgelerin şartlarına bağlı olarak araştırmalar yapılmıştır. Literatürde araştırmalarda göçü etkileyen çeşitli faktörlerin incelendiği gözlemlenmiştir. Bu çalışmanın amacı Türkiye'nin 81 ili arasındaki göçü 2013-2023 yıllarında ekonomi, sağlık, kültür ve şehrin altyapısını temsil edebilecek değişkenlerle analiz etmektir. Değişkenlerin tamamı TÜİK'ten elde edilmiştir. Elde edilen değişkenlere ait veri setleri çalışma kapsamında hazırlanan arayüz ile farklı formatlardan standart formata dönüştürülmüş ve hatasız bir şekilde “Şehir-Yıl” anahtar kabul edilerek birleştirilmiştir. Final veri seti Panel Veri Analizi, Yerçekimi Modeli ve Sosyal Ağ Analizi yöntemlerinde kullanılmak için hazırlanmıştır. Panel veri analizinde iki farklı bağımlı değişken ile iki model kurulmuştur. Çıktılarda paralellikler saptanmıştır. Yerçekimi modelinde bağımlı değişken yıllık göç akışıdır. Sonuçlar iller arası mesafenin fazla olmasının göçü negatif etkilediğini göstermiştir. Sosyal ağ analizinde de Türkiye 5 havzaya ayrılmıştır. Böylece bu tez çalışması bölgesel kalkınma politikaları ve şehir planlaması için veri temelli bir rehber sunmaktadır.

Giriş

İç göç literatürü çoğunlukla temel ekonomik/demografik göstergelere odaklanırken; sağlık hizmetleri, kültür-sanat olanakları ve emlak piyasası gibi unsurların göç kararında önemli olmasına rağmen çoğu çalışmada ihmal edildiği görülmektedir. Bu tez, sağlık-kültür-ekonomi değişkenlerini bütüncül biçimde aynı çerçevede ele alır ve 2013-2023 gibi güncel ve uzun bir dönem ile katkı sunmayı hedefler. Özgün yönlerden biri: panel veri ve yerçekimi modelini birlikte kullanarak göçün hem zaman içindeki dönüşümünü hem de mekânsal etkileşimini eş zamanlı incelemek.



Veri

Kapsam: 81 il × 2013-2023; göç ve sosyo-ekonomik göstergeler. (Değişkenlerin tamamı TÜİK'ten elde edilmiştir.)

Tablo 1: Çalışmaya Dahil Edilmiş Tüm Değişkenler

Kategori	Değişkenler	Ölçüm Birimi
Bağımlı Değişken	Aldığı Göç, Verdiği Göç	Kişi Sayısı
Ekonomi	GSYH, Girişim Sayısı, İthalat, İhracat	TL / Adet / USD
Sosyal/Demografi	Nüfus Yoğunluğu, Evlilik, Boşanma Hızı, Hane halkı Büyüklüğü	Oran / Sayı
Sağlık	Hekim Sayısı, Hastane Yatak Sayısı	1000 Kişi Başına / Sayı
Kültür/Yaşam	Sinema, Tiyatro, Kütüphane, Otomobil Sayısı	Salon Sayısı / Sayı
Konut	Konut Satışı, Yapı Belgesi (Daire Sayısı)	Adet

Veri mühendisliği: TÜİK ham verileri farklı formatlardan, geliştirilen arayüz ile standart forma dönüştürülmüş; 19 modül ortak “**Şehir-Yıl**” anahtarında birleştirilerek Dengeli Panel (Balanced Panel) oluşturulmuştur.

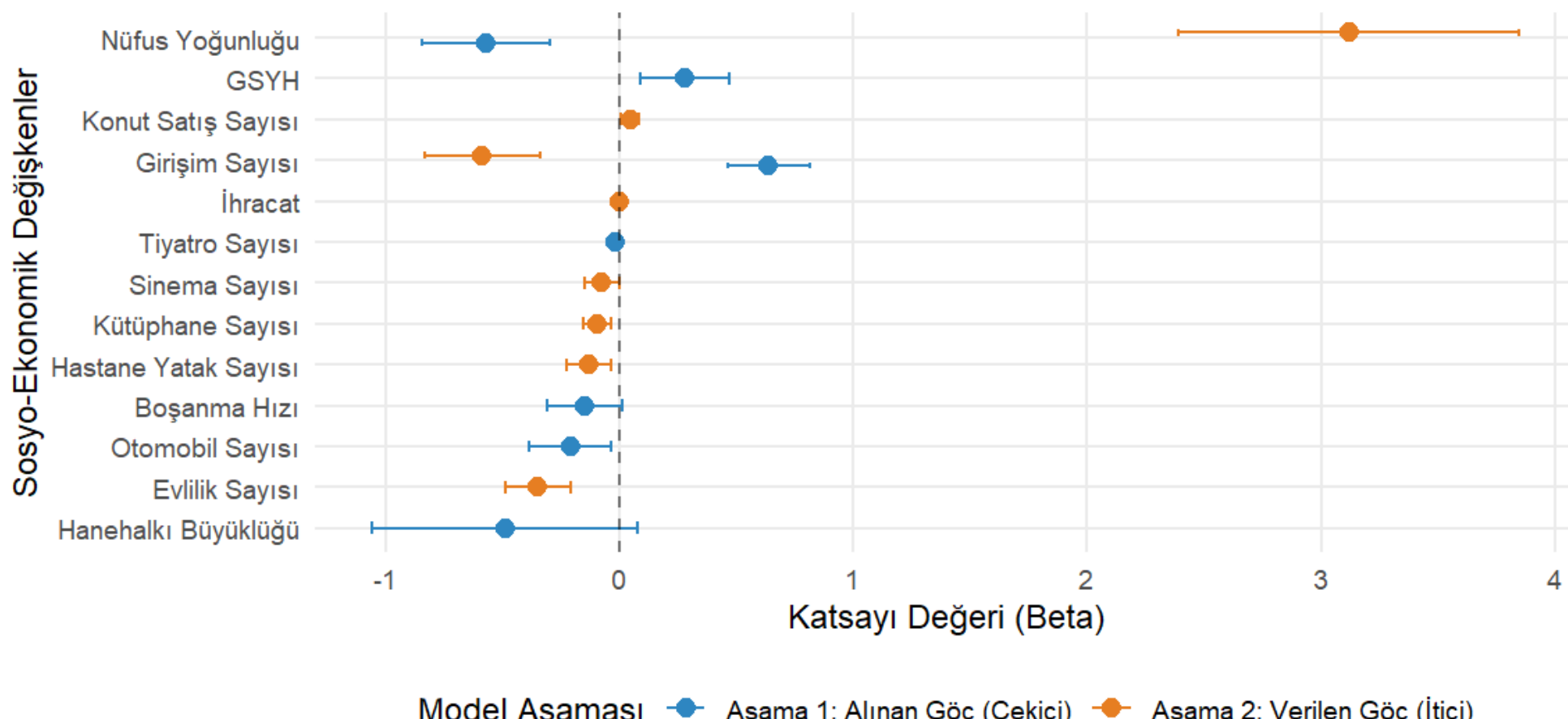


Yöntem

1) Panel Veri Analizi (2 aşama, sabit etkiler):

- Aşama 1:** Alınan göç = çekici faktörler;
- Aşama 2:** Verilen göç = itici faktörler.
- Model seçimi/varsayım testleri:** Hausman testi sabit etkileri destekler; Breusch-Pagan değişen varyans sorununa işaret eder; bu nedenle Driscoll-Kraay dirençli tahminci ile nihai sonuçlara ulaşılır.

Şehirler Arası Göçün Belirleyicileri: Karşılaştırmalı Analiz
Logaritmik Esneklik Katsayıları ve %95 Güven Aralıkları



2) Yerçekimi (Gravity) Modeli – PPML:

Mesafe, illerin coğrafi merkezleriyle Haversine yöntemiyle hesaplanır.

Tahmin yöntemi: literatürde “altın standart” kabul edilen PPML; sıfır akışları dışlamadan ve heteroskedastisite sorununu aşarak tutarlı tahmin üretir; kaynak/varış/yıl sabit etkileri kullanılır.

Tablo 2: Yerçekimi Modelindeki İstatistiksel Olarak Anlamlı Değişkenler

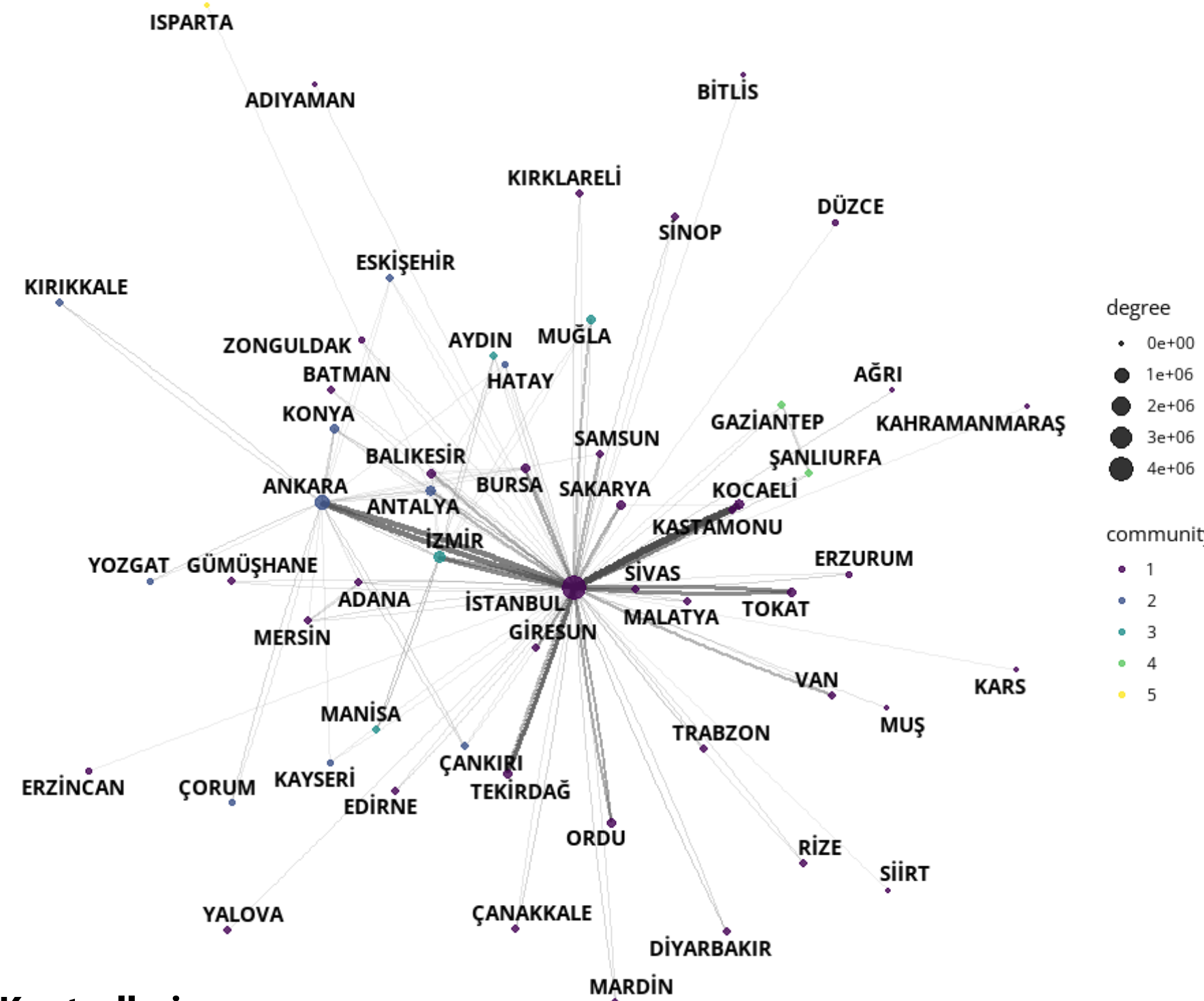
Bağımsız Değişkenler	Katsayı (Estimate)	Standart Hata	z-değeri	p-değeri
Mesafe (ln_dist)	-0.6865***	0.0309	-22.20	< 0.001
GSYH Farkı (diff_gsyh)	-0.1557***	0.0403	-3.86	< 0.001
Hekim Sayısı Farkı (diff_hekim)	-0.2492***	0.0694	-3.59	< 0.001
Girişim Sayısı Farkı (diff_girisim)	0.5838***	0.1225	4.76	< 0.001

3) Sosyal Ağ Analizi – Walktrap:

İller arası göç ağında Walktrap ile 5 havza/topluluk tespit edilir; ağ görselinde renkler bu kümeleri gösterir.

Türkiye İller Arası Göç Ağı ve Sosyolojik Kümeler

En güçlü %2'lik akışlar ve Walktrap topluluk tespiti



4) Sağlıklik Kontrolleri:

LOYO (Leave-One-Year-Out) ile katsayı istikrarı test edilmiştir; katsayıların dar bantta kaldığı ve sonucun belirli bir yıla bağımlı olmadığı gösterilmiştir.

Sonuç

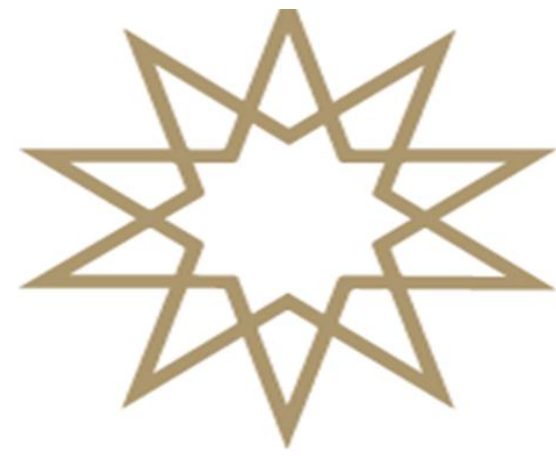
Bu çalışmada, Türkiye'nin 81 ili arasındaki 2013-2023 dönemi iç göç hareketleri üç farklı metodolojik yaklaşımla kapsamlı olarak incelenmiştir. Panel veri analizi sonuçları, göç hareketliliğinin %94'ten fazlasının illerin yapısal farklarıyla açıklandığını, zaman içi dalgalanmaların minimal olduğunu göstermiştir. Nüfus yoğunluğu en baskın itici faktör olarak tespit edilirken, girişimcilik kapasitesi hem çekici hem tutucu bir güç olarak öne çıkmıştır. Yerçekimi modeli, mesafenin hala göç önünde önemli bir engel olduğunu ve göçün belirli merkezlerde yapısal olarak kümelandığını ortaya koymuştur. Sosyal ağ analizi, Türkiye'nin 5 entegre göç havzasından oluştuğunu ve Marmara aksının ulusal çekim merkezi olduğunu göstermiştir.

Öneriler

Elde edilen bulgular ışığında, bölgesel kalkınma politikalarının 81 il bazında değil, tespit edilen 5 havza bazında kurgulanması önerilmektedir. Göç veren illerde mikro-girişimcilik ve iş kurma destekleri önceliklendirilerek bölgelerin ekonomik çekim gücü artırılabilir. Kütüphane, sinema ve hastane gibi yaşam kalitesini artıran sosyal ve fiziksel altyapı yatırımları, sadece ekonomik teşvikler değil, tutucu güçler olarak planlanmalıdır. Büyükşehirlerde nüfus yoğunluğunu azaltmak için tespit edilen havzalar içindeki alt merkezlere nüfus kaydırma stratejileri izlenmelidir. Mesafenin engelleyici etkisini azaltmak için uzak bölgeler arası lojistik ve ekonomik entegrasyon güçlendirilmeli, dijital altyapı ile sosyal mesafe azaltılmalıdır. Bu tez çalışması, dengeli bölgesel kalkınma ve sürdürülebilir şehir planlaması için veri temelli bir rehber sunmaktadır.

KAYNAKÇA

- [1] Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, 23(4), 1249-1274.
- [2] Poot, J. et al. (2016). The gravity model of migration: the successful comeback of an ageing superstar. *Investigaciones Regionales*, 36, 63-86.
- [3] Arif, I. (2022). Educational attainment, corruption, and migration: An empirical analysis from a gravity model. *Economic Modelling*, 110.
- [4] Pitoski, D. et al. (2024). The complex network patterns of human migration. *Applied Network Science*, 9(1).
- [5] TÜİK (2025). *Nüfus İstatistikleri Portalı*. <https://nup.tuik.gov.tr/>
- [6] TÜİK (2025). *Veri Portalı - MEDAS*. <https://data.tuik.gov.tr/>

2025-2026 Güz Yarıyılı
İSTATİSTİK BÖLÜMÜ

FUTBOLCULARIN PİYASA DEĞERİ DEĞİŞİMİNDE SAHA İÇİ PERFORMANS, TAKIM VE LİG FAKTÖRLERİNİN ANALİZİ

Kemal Hakan Kalkuz 20023023

Danışman: Dr. Öğr. Üyesi Tugay KARADAĞ

ÖZET

2024-2025 futbol sezonu içerisinde Avrupa'nın 5 büyük liginde mücadele eden 2395 futbolcuya ait piyasa değerlerindeki değişimler, bahsi geçen sezon içerisindeki saha içi performans metrikleri, takım başarısı ve lig kalitesi gibi faktörlerle analiz edilmiştir. Kaleci, Defans, Orta Saha ve Forvet mevki grupları için ayrı olarak yapılan 4 adet Çoklu Doğrusal Regresyon Analizi sonucunda, 277 değişken içerisinde 16 değişkenin piyasa değeri değişimi üzerinde istatistiksel olarak anlamlı etkisi olduğu gözlemlenmiştir.

GİRİŞ

Futbol yıllar içerisinde sahadaki rekabetin ötesine geçerek transfer bedellerinin, maaşların ve kulüp bütçeleri toplamının milyar dolarlarla ifade edildiği devasa bir ekonomiye dönüşmüştür. Bu ekosistemin sürdürülebilirliği açısından en kritik unsurlardan biri de futbolcuların transfer sürecinde ödenecek bonservis bedelleridir.

AMAÇ

Bu çalışmanın temel amacı, 2024-2025 sezonu resmî Avrupa Futbol Federasyonları Birliği (UEFA: Union of European Football Associations) sıralamalarına göre Avrupa'nın 5 büyük liginde (Almanya, Fransa, İngiltere, İspanya ve İtalya) mücadele eden 96 adet futbol takımının bünyesinde yer alan 2395 futbolcunun sezon başı ve sezon sonu piyasa değeri değişimlerinde en büyük etkiye sahip değişkenleri tespit etmektir.

VERİ SETİ

Çalışmada kullanılan veri seti, FBref ve Transfermarkt platformları üzerinden alınmıştır. 2024-2025 futbol sezonunda Avrupa'nın 5 büyük liginde forma giyen oyunculara ait istatistikleri içeren FBref temelli veri seti, Kaggle üzerinden indirilmiştir. Ardından bahsi geçen oyuncular için belirtilen futbol sezonu başlangıcı ve sonu dönemlerine ait piyasa değerleri Transfermarkt platformu üzerinden çekilmiştir. Elde edilen 2 veri seti, futbolcuların kimlik değişkenlerine (ad, takım, yaş gibi) sadık kalınarak birleştirilmiştir. Toplam 2395 adet futbolcu için saha içi istatistikleri ve piyasa değer değişimlerini temsil eden 277 adet değişken bulunmaktadır. Regresyon modeli kurulum sürecinde Multicollinearity (Çoklu Doğrusal Bağlantı) sorunları yaşamamak adına birbirleri ile yüksek ilişki gösteren değişkenler elenerek 95 adet analiz sürecine dahil edilmiştir. Veri seti içerisinde bulunan ve sezon içi piyasa değer değişimini temsil eden «fark» değişkeni, bağımlı değişken olarak seçilmiştir.

player	pos	squad	comp	age	mp	starts	min
Arda Guler	MF	FW Real Madrid	es La Liga	19	28	14	1250
Erling Haaland	FW	Manchester City	eng Premier League	24	31	31	2736
Gianluigi Donnarumma	GK	Paris S-G	fr Ligue 1	25	24	24	2091
Jude Bellingham	MF	Real Madrid	es La Liga	21	31	29	2488
Kevin De Bruyne	MF	FW Manchester City	eng Premier League	33	28	19	1702
Kylian Mbappe	FW	Real Madrid	es La Liga	25	34	34	2907
Lamine Yamal	FW	Barcelona	es La Liga	17	35	31	2856
Marco Asensio	MF	Aston Villa & Paris S-G	eng Premier League	28	25	17	1361
Ousmane Dembele	FW	Paris S-G	fr Ligue 1	27	29	20	1730
William Saliba	DF	Arsenal	eng Premier League	23	35	35	3039

ÇOKLU DOĞRUSAL REGRESYON ANALİZİ

Çoklu doğrusal regresyon analizi, bir bağımlı değişken ile iki veya daha fazla bağımsız değişken arasındaki doğrusal ilişkiyi incelemek ve bu ilişkiden yararlanarak tahmin modelleri oluşturmak için kullanılan istatistiksel bir yöntemdir. Bu çalışma kapsamında, futbolcuların piyasa değerini etkileyen performans metrikleri, takım ve lig faktörleri arasındaki ilişki bu yöntemle modellenmiştir. Analizin temel amacı, bağımsız değişkenlerdeki bir birimlik değişimin bağımlı değişken üzerinde yarattığı etkini miktarını ve yönünü (pozitif/negatif) belirlemektir.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

y = Bağımlı Değişken (Piyasa Değeri)

β_0 = Sabit Terimi

$\beta_1 \dots \beta_k$ = Bağımsız değişkenlerin regresyon katsayısı

$X_1 \dots X_k$ = Bağımsız değişkenler (age, 90s, takım_uefa_katsayısı vb.)

ϵ = Hata Terimi

SONUÇ

Analize dahil edilen 95 adet değişken ilk olarak futbolcuların mevkillerine dikkat edilerek gruplandırılmıştır. Kaleci, Defans, Orta Saha ve Forvet grupları için gruplandırılan her değişken modellere dahil edilerek test edilmiştir. Analiz sürecinde, futbolcuların toplam oynama sürelerinin 90 dakikaya oranlanması ile elde edilen “90s” değişkenine ait verileri en az 7 olan futbolcular örnekleme dahil edilmiştir. Modellerin kurulması ile birlikte 16 adet değişkenin futbolcuların piyasa değerleri üzerinde istatistiksel olarak anlamlı sonuç verdiği gözlemlenmiştir.

Kısaltma	Açıklama
90s	90 dakikaya oranlanmış toplam oynama süresi
age	Oyuncunun yaşı
ast	Gol ile sonuçlanan pas sayısı
blocks	Rakip şut veya paslarını engelleme sayısı
crdy	Sezon içinde alınan toplam sarı kart sayısı
g-pk	Penaltı vuruşları hariç kaydedilen toplam gol sayısı
g/sot	İsabetli şut başına kaydedilen gol sayısı
np:g-xg	Penaltı vuruşları hariç, gol ile beklenen gol değeri farkı
np:xg	Penaltı vuruşları hariç, toplam beklenen gol değeri
onxg	Oyuncu sahadayken takımı tarafından üretilen toplam beklenen gol değeri
onxga	Oyuncu sahadayken takımının kalesinde karşılaştığı toplam beklenen gol değeri
sot	Toplam isabetli şut sayısı
takim_uefa_katsayısı	Kulübün resmî UEFA başarı katsayısı
thr	Kalecinin eli ile topu oyuna sokma sayısı
touches	Oyuncunun top ile toplam buluşma ve temas sayısı
xag	Oyuncunun verdiği pas sonucu takım arkadaşının şutunun gol olma olasılığı

Kaleciler için elde edilen sonuçlara göre, age ve thr değişkenleri piyasa değeri üzerinde değişime sebep olmaktadır. Bu değişkenler ile kurulan regresyon modeli, kalecilerin piyasa değeri değişimini %17,14 oranında açıklamaktadır.

$$fark=1,883,986.06-318,176.87(age)+44,449.61(thr)+ \epsilon$$

MODEL ÖZETİ:
R-squared: 0.3228
Adjusted R-squared: 0.1714
F-statistic: 2.13
Prob (F-statistic): 0.009774
AIC: 3491.14
BIC: 3544.22

Defans oyuncularını için elde edilen sonuçlarda, «age», «onxg», «takim_uefa_katsayısı», «sot», «ast», «onxga» ve «blocks» değişkenlerinin piyasa değer değişimi üzerinde anlamlı etkilere sahip olduğu gözlemlenmiştir. Bu değişkenler ile kurulan regresyon modeli, defans oyuncularını için piyasa değer değişimini %19,39 oranında açıklamaktadır.

$$fark=6,129,618.62-438,236.40(age)+174,528.74(onxg)-34,454.96(uefa_takim_katsayısı)-544,733.21(sot)+554,242.2(ast)-84,265.59(onxga)-81,243.90(blocks)+ \epsilon$$

MODEL ÖZETİ:
R-squared: 0.2584
Adjusted R-squared: 0.1939
F-statistic: 4.43
Prob (F-statistic): 0.000000
AIC: 17352.46
BIC: 17509.42

Orta Saha oyuncularını için, «age», «takim_uefa_katsayısı», «onxg», «np:g-xg», «onxga», «crdy», «g/sot» ve «g-pk» değişkenleri anlamlı sonuçlar vermiştir. Bu değişkenler ile kurulan model orta saha oyuncularının piyasa değer değişimini %25,58 oranında açıklamaktadır.

$$fark=21,217,929.52-539,998.15(age)-71,691.82(takim_uefa_katsayısı)+281,982.72(onxg)+671,585.05(np:g-xg)-168,070.68(onxga)-328,804.83(crdy)-3,993,526.14(g/sot)+461,422.18(g-pk)+ \epsilon$$

MODEL ÖZETİ:
R-squared: 0.3187
Adjusted R-squared: 0.2558
F-statistic: 5.87
Prob (F-statistic): 0.000000
AIC: 14189.85
BIC: 14334.87

Son olarak forvet oyuncularını için, «age», «takim_uefa_katsayısı», «90s», «xag», «np:xg», «np:g-xg», «g-pk» ve «touches» değişkenleri anlamlı sonuçlar göstermiştir. Bu değişkenler ile kurulan model forvet oyuncularını için piyasa değer değişimini %31,73 oranında açıklamaktadır.

$$fark=23,546,803.12-744,903.77(age)-115,561.02(takim_uefa_katsayısı)-821,085.20(90s)+1,881,900.98(xag)+7,743,416.94(np:xg)+7,569,391.43(np:g-xg)-6,513,582.80(g-pk)+21,688.28(touches)+ \epsilon$$

MODEL ÖZETİ:
R-squared: 0.3927
Adjusted R-squared: 0.3173
F-statistic: 5.21
Prob (F-statistic): 0.000000
AIC: 10740.31
BIC: 10870.98

TARTIŞMA

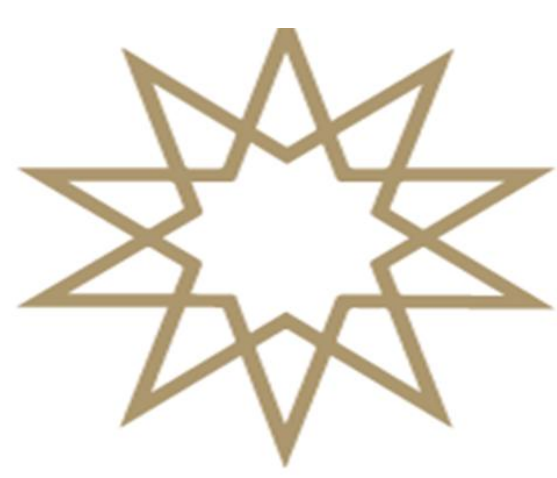
Elde edilen bulgular ışığında, futbolcuların piyasa değerlerinin tahminlerden uzak, onlarca farklı faktöre bağlı olarak hesaplanan karmaşık değerler olduğu görülmüştür. Mustafa Demir, futbolcular üzerine yaptığı çalışmasında piyasa değeri ve transfer ücretleri arasındaki değişkenliğe sebep olan saha dışı faktörlere dikkat çekerek, modellerin başarısını artırmak için bu etkenlerin önemini vurgulamıştır^[1]. Bu çalışmada kurulan Çoklu Doğrusal Regresyon modelleri, Avrupa'nın 5 büyük ligindeki futbolcuların piyasa değeri değişimlerini mevki bazlı olarak yaklaşık %17,14 ile %31,73 arasında değişen oranlarda açıklamaktadır. Benzer bir çalışmada bulunan Yu Sun ve Kepenk Gu, makine öğrenmesi yöntemleriyle %99 gibi oldukça yüksek bir piyasa değeri tahmin başarısı elde etmişlerdir^[2]. Bizim çalışmamız ile arasında bulunan temel fark ise amaç ve yöntemdir. Yu Sun ve Kepeng Gu piyasa değerini tahmin etmek için makine öğrenimi üzerine odaklanmışken, biz bu çalışmada Çoklu Doğrusal Regresyon Analizi yöntemi kullanarak saha içi verilerinin, takım ve lig bilgilerinin piyasa değeri değişimi üzerinde ne yönde ve ne büyüklükte etkiye sahip olduğunu ortaya çıkardık. Yaptığımız çalışma sonucunda futbolcunun mücadele ettiği ligin UEFA başarı katsayısı, mevki fark etmeksizin piyasa değeri üzerinde anlamlı bir etkisi olmadığı görülmüştür. Literatür araştırması sürecinde karşılaşılan bir diğer çalışmada Geurts, bizim sonuçlarımızdan farklı olarak futbolcuların üst düzey liglerde mücadele etmesinin piyasa değerini artıran bir faktör olduğunu tespit etmiştir^[3].

KAYNAKÇA

[1] Demir, M. (2019). *Profesyonel futbolcu piyasa değeri ve transfer ücreti arasındaki değişkenliği sağlayan etkenlerin değerlendirilmesi*. Marmara Üniversitesi Sağlık Bilimleri Enstitüsü.

[2] Sun, Y., & Gu, K. (2024). Prediction of Football Players' Value in the Transfer Market of Well-known European Leagues based on FIFA 19 and Real-world Data. *International Arab Journal of Information Technology (IAJIT)*.

[3] Geurts. (2016). *Football players' transfer price determination based on performance in the Big 5 European leagues*. Universidade Nova de Lisboa.



YTU

FEN EDEBİYAT FAKULTESİ

İSTATİSTİK BÖLÜMÜ

Beşiktaş'taki Hava Kalitesini Etkileyen Faktörlerin Makine Öğrenmesiyle İncelenmesi

SÜEDA USTAMEHMETOĞLU 20023056

Danışman: Dr. Öğr. Üyesi COŞKUN PARİM

ÖZET

Hava kirliliği, özellikle büyük şehirlerde insan sağlığı ve çevre üzerinde ciddi tehditler oluşturmaktadır. PM₁₀, solunum yollarına kadar ulaşabilen yapısı nedeniyle hava kalitesinin en kritik göstergelerinden biridir ve meteorolojik koşullar ile kentsel emisyonlardan doğrudan etkilenmektedir. Bu çalışma, PM₁₀ seviyelerini geniş bir kirlletici ve meteorolojik değişken seti kullanarak modellemeyi ve geleceğe yönelik tahmin gücünü artırmayı amaçlamaktadır. Regresyon tabanlı makine öğrenmesi modelleri ve hiperparametre optimizasyonu ile PM₁₀ tahmin doğruluğu yükseltilmiş, sonuçlar hava kalitesi yönetimi için güçlü bir karar destek altyapısı sunduğunu göstermiştir.

VERİ SETİ

Bu çalışmada, Beşiktaş ilçesinin hava kalitesini etkileyen faktörler hava kalitesi ve meteorolojik veriler kullanılarak incelenmiştir. Analiz, 2019 – 2024 tarih aralığındaki günlük verileri kapsamaktadır. Hava kalitesi verileri T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Sürekli İzleme Merkezi, meteorolojik veriler ise “Open-Meteo Tarihsel Hava Durumu API” kaynağı aracılığıyla temin edilmiştir.

Çalışmada PM₁₀, PM_{2.5}, SO₂, CO, NO₂, NO_x, NO ve O₃ kirlleticileri ile sıcaklık, yağış, nem, rüzgâr, güneşlenme, radyasyon ve bulutluluk değişkenleri modele dahil edilmiştir.

VERİ BİRLEŞİMİ VE VERİ DOLDURMA

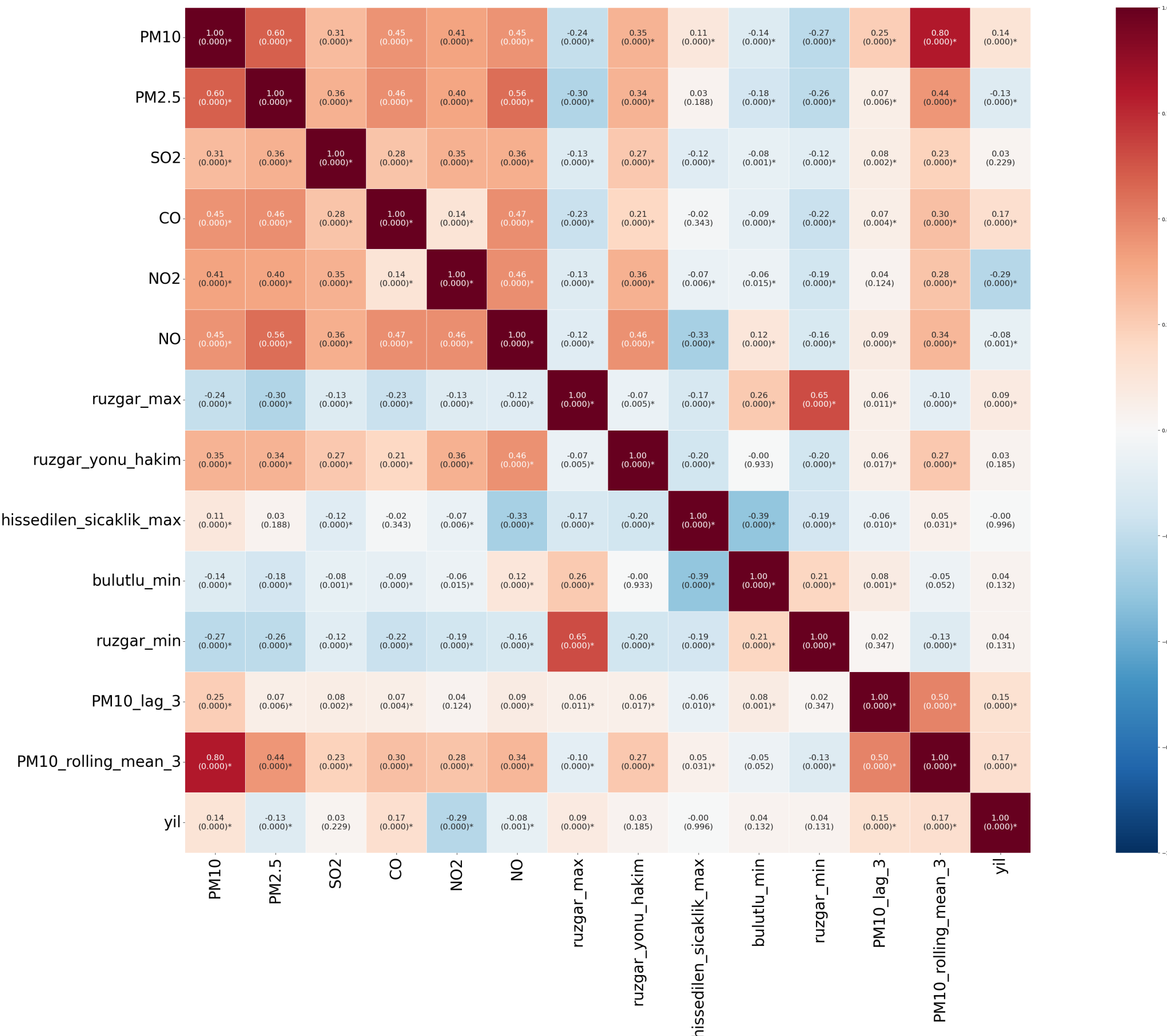
Hava kalitesi ve meteorolojik verileriyle oluşan veri setini Tarih sütunu üzerinden birleştirilmiştir. PM₁₀ dışındaki değişkenlerdeki eksik değerler, örneklem kaybını önlemek amacıyla K-En Yakın Komşu (KNN) yöntemi ile doldurulmuştur. Bu yöntemde, Öklid mesafesine göre en yakın 5 komşu kullanılarak eksik değerler tahmin edilmiş, yalnızca sayısal değişkenler işleme dahil edilmiştir.

VERİ ÖLÇEKLENDİRME

Değişkenler, çarpıklık katsayılarına göre otomatik olarak dönüştürülmüştür. Yüksek çarpıklığa sahip değişkenler Yeo-Johnson güç dönüşümü, normale yakın dağılıma sahip olanlar Z-skor standardizasyonu, farklı ölçeklerdeki değişkenler ise Min-Max ölçekleme ile düzenlenmiştir.

VIF kontrolü yapılarak değişkenlerin 5 sınır değerinin altında değerler aldığı görüldü. Bu nedenle çoklu doğrusal bağlantı sorunu bitirilmiş oldu.

KORELASYON



Çoklu bağlantı sorunu yaratan değişkenlerin VIF analizi ve özellik seçimi yöntemleriyle elenmesi sonucunda oluşturulan nihai veri setine ait korelasyon matrisi incelendiğinde; hedef değişken PM₁₀ ile en güçlü ilişkinin 0.80 katsayısı ile tarafımızca türetilen 3 Günlük Hareketli Ortalama değişkeni arasında olduğu görülmektedir. Bu bulgu, hava kirliliğinin anlık meteorolojik olaylardan ziyade geçmiş günlerin birikimli trendlerine (zamansal bağımlılık) dayandığını kanıtlarken; PM_{2.5} ile tespit edilen 0.60 düzeyindeki pozitif ilişki ortak emisyon kaynaklarına, rüzgar hızı ile gözlemlenen negatif korelasyon ise atmosferik faktörlerin kirliliği azaltıcı (ventilasyon) etkisine işaret ederek modelin fiziksel gerçeklikle uyumlu, sade ve kararlı bir yapıya kavuştuğunu göstermektedir.

YÖNTEM

Zamansal bağımlılığı korumak amacıyla veri seti kronolojik olarak bölünmüş; ilk %80 eğitim/validasyon, son %20 test için kullanılmıştır. Hiperparametre optimizasyonunda zaman serisi çapraz doğrulama (genişleyen pencere) uygulanarak bilgi sızıntısı önlenmiştir. Tahmin modeli olarak XGBoost tercih edilmiş, aşırı öğrenmeyi engellemek için düşük ağaç derinliği, düzenleştirmeye ve düşük öğrenme oranları kullanılmıştır. En iyi model, R² metriğine göre seçilerek test verisi üzerinde değerlendirilmiştir.

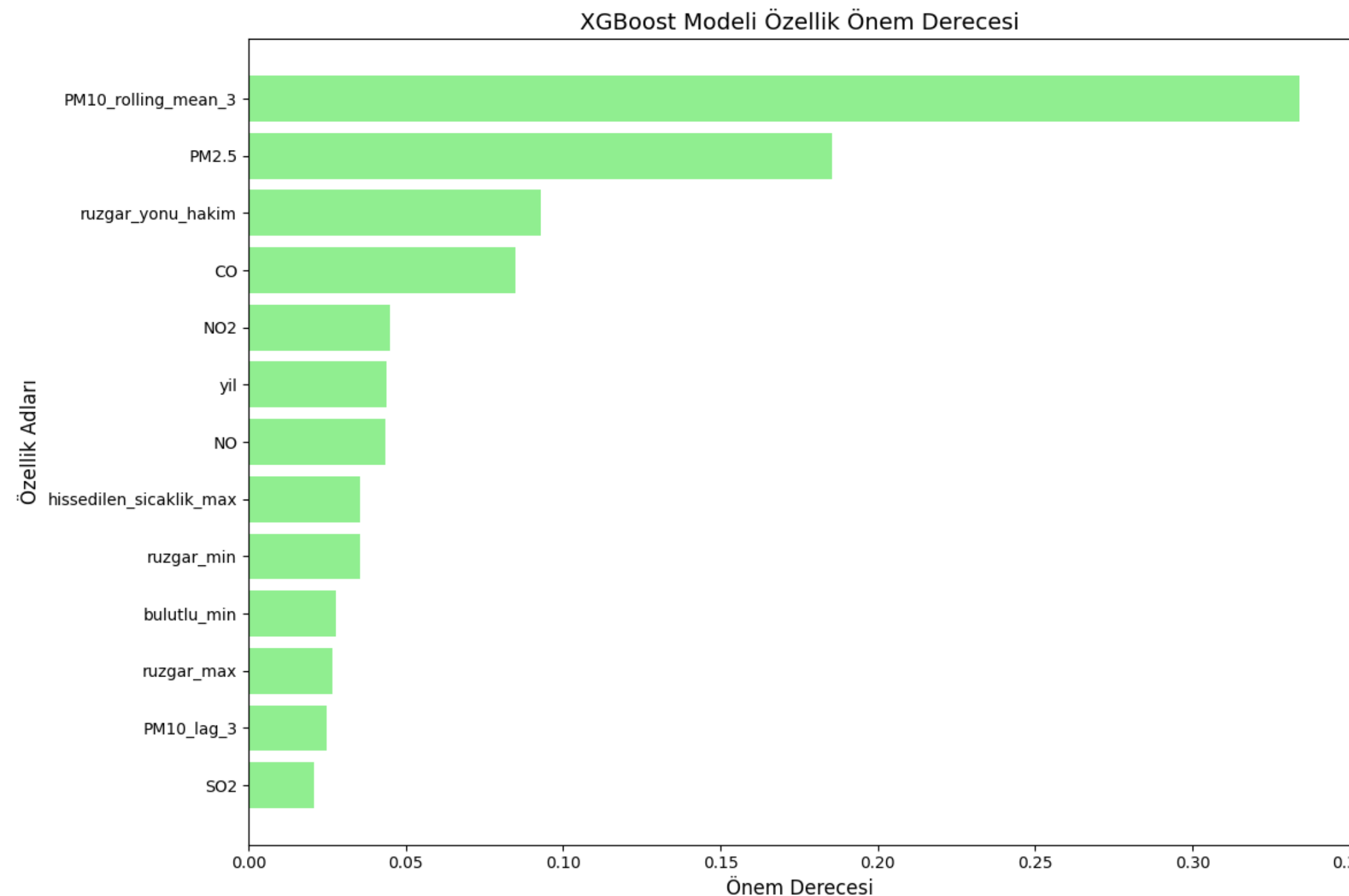
	Eğitim			Test		
	MSE	RMSE	R ²	MSE	RMSE	R ²
Karar Ağacı	0.2222	0.4714	0.7607	0.3860	0.6213	0.6178
K-En Yakın Komşu	0.1766	0.4202	0.8098	0.5274	0.7262	0.4777
Rasgele Orman	0.0314	0.1773	0.9661	0.3366	0.5802	0.6667
XGBoost	0.1369	0.3700	0.8600	0.2609	0.5108	0.7501

Geliştirilen modellerin performansları karşılaştırıldığında, XGBoost algoritmasının doğruluk ve genelleme yeteneği açısından en başarılı sonuçları verdiği görülmektedir. Rasgele Orman modeli eğitim setinde yüksek başarı göstermesine rağmen test setinde belirgin performans kaybı yaşayarak aşırı öğrenme (overfitting) eğilimi sergilemiştir. Buna karşılık XGBoost, test setinde elde ettiği 0.75 R² ve 0.51 RMSE değerleriyle en yüksek tahmin gücüne ulaşmıştır. Eğitim (R² = 0.86) ve test performansları arasındaki tutarlılık, modelin gürültüden arındırılmış ve yüksek genelleme yeteneğine sahip olduğunu göstermektedir.

SONUÇ

Özellik önem analizi, PM₁₀ kirliliğinin oluşumunda en baskın belirleyicilerin geçmiş kirlilik trendleri (PM₁₀ rolling mean) ve PM_{2.5} seviyeleri olduğunu göstermektedir. Meteorolojik değişkenler arasında hâkim rüzgâr yönünün öne çıkması, Beşiktaş bölgesindeki PM₁₀ kirliliğinin yalnızca yerel emisyonlardan değil, aynı zamanda atmosferik taşınım süreçlerinden de önemli ölçüde etkilendiğine işaret etmektedir.

Bu bulgular doğrultusunda, çalışmada uygulanan özellik mühendisliği yaklaşımı ve optimize edilmiş XGBoost modeli, karmaşık ve kaotik yapıya sahip atmosferik süreçlerin yüksek doğrulukla modellenebileceğini ortaya koymuştur. Elde edilen sonuçlar, geliştirilen modelin yerel yönetimler için güvenilir bir karar destek ve erken uyarı sistemi olarak kullanıma potansiyeline sahip olduğunu göstermektedir.



KAYNAKÇA

- [1] Open-Meteo, *Meteorolojik veri seti*, <https://open-meteo.com/>.
- [2] T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, *Sürekli İzleme Merkezi İstasyon Verileri*, https://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew.
- [3] Scikit-learn developers, “Tree-based models,” scikit-learn. <https://scikit-learn.org/stable/modules/tree.html> (Erişim Tarihi: Ekim 8, 2024).



YTÜ

FEN EDEBİYAT FAKÜLTESİ

İSTATİSTİK BÖLÜMÜ

IMDb Dizilerinde Bölüm Sayısı ile İzleyici Puanı Arasındaki İlişkinin İstatistiksel Olarak İncelenmesi

Muhammed Fazıl ÇOBAN 19023070

Danışman: Doç. Dr. Atif Ahmet EVREN

ÖZET

Bu çalışmada IMDb platformunda yer alan dizilerde bölüm sayısı ile izleyici puanı arasındaki ilişki istatistiksel yöntemler kullanılarak incelenmiştir. Analizler 1980–2025 yılları arasında yayınlanmış 86.221 dizi üzerinden gerçekleştirilmiştir. Bulgular, bölüm sayısının izleyici puanları üzerindeki etkisinin sınırlı olduğunu, tür faktörünün ise daha belirleyici olduğunu göstermektedir.

AMAÇ

Bu çalışmanın amacı, IMDb dizilerinde bölüm sayısı ile izleyici puanı arasında istatistiksel olarak anlamlı bir ilişki olup olmadığını incelemektir. Özellikle uzun soluklu dizilerin izleyici değerlendirmeleri üzerindeki etkisi analiz edilmiştir.

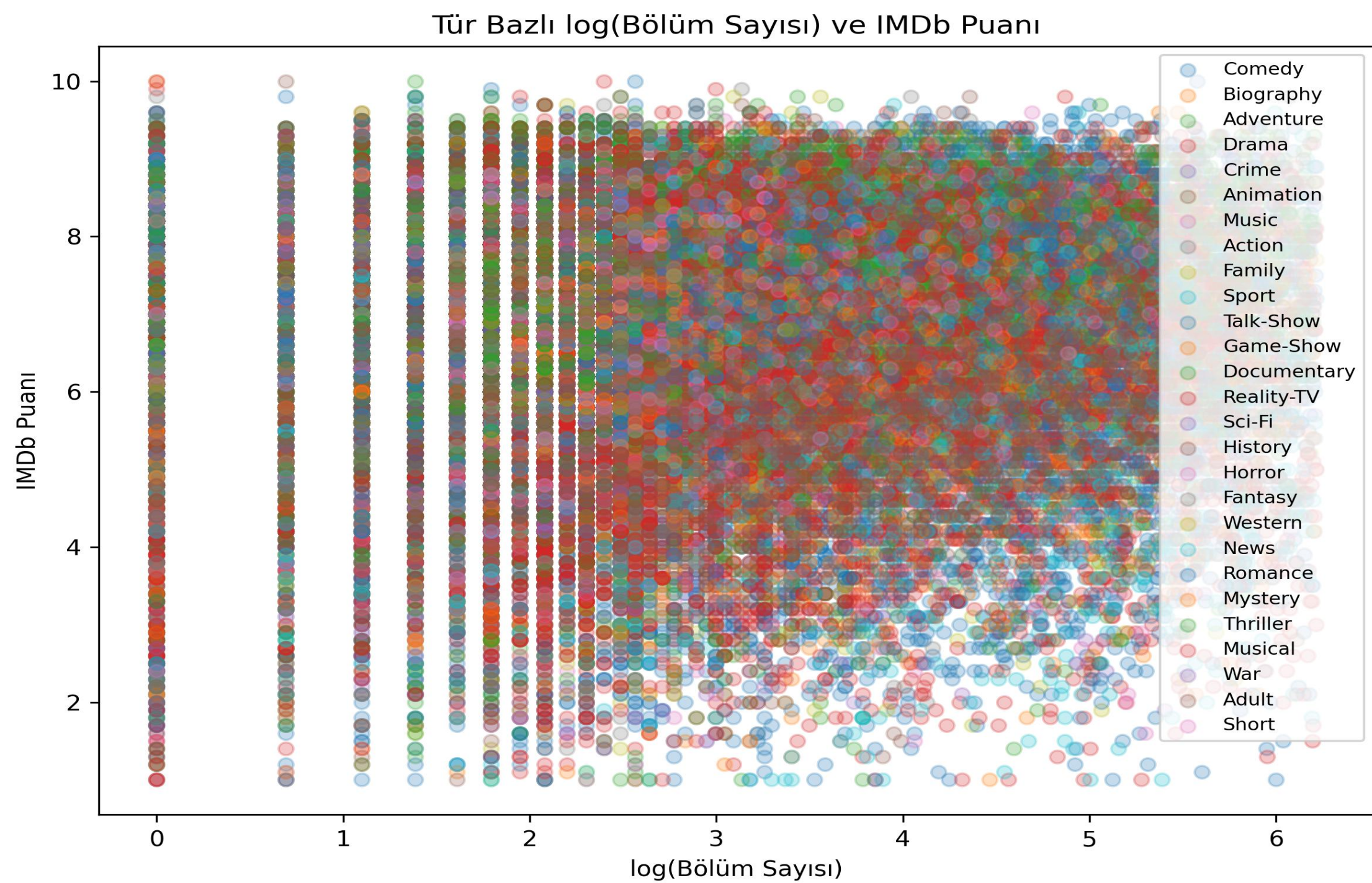
VERİ SETİ

- IMDb dizileri (1980–2025)
- Toplam 86.221 dizi
- Bağımlı değişken: İzleyici puanı (regresyon modelinde kullanılmıştır)
- Bağımsız değişken: Bölüm sayısı (logaritmik dönüşüm sonrası regresyonda kullanılmıştır)

Title	Rating	Episodes	Year	Genre
Bracken	7.6	12	1980	Drama
Wheel of Fortune	6.7	105	1981	Game-Show
Los gozos y las sombras	8.4	13	1982	Drama
The 1/2 Hour Comedy Hour	6.2	1	1983	Comedy
The Steam Video Company	7.5	6	1984	Comedy
227	6.8	116	1985	Comedy
Blackadder II	8.6	6	1986	Comedy
Varuhuset	6.2	60	1987	Drama
After Henry	7.0	38	1988	Comedy
Saturday Night Clive	7.6	46	1989	Talk-Show
Fredrikssons fabrikk	6.5	18	1990	Comedy
Farmacia de guardia	6.6	172	1991	Comedy
Bony	7.7	13	1992	Crime
Donahue/Pozner	6.9	1	1993	Talk-Show
Orson & Olivia	8.1	26	1994	Animation
American Cinema	7.6	13	1995	Documentary
London Bridge	8.1	237	1996	Drama
413 Hope St.	6.1	10	1997	Drama
Style & Substance	7.9	13	1998	Comedy
Clube dos Campeões	7.6	26	1999	Comedy
Beast	8.1	12	2000	Comedy

KORELASYON ANALİZİ

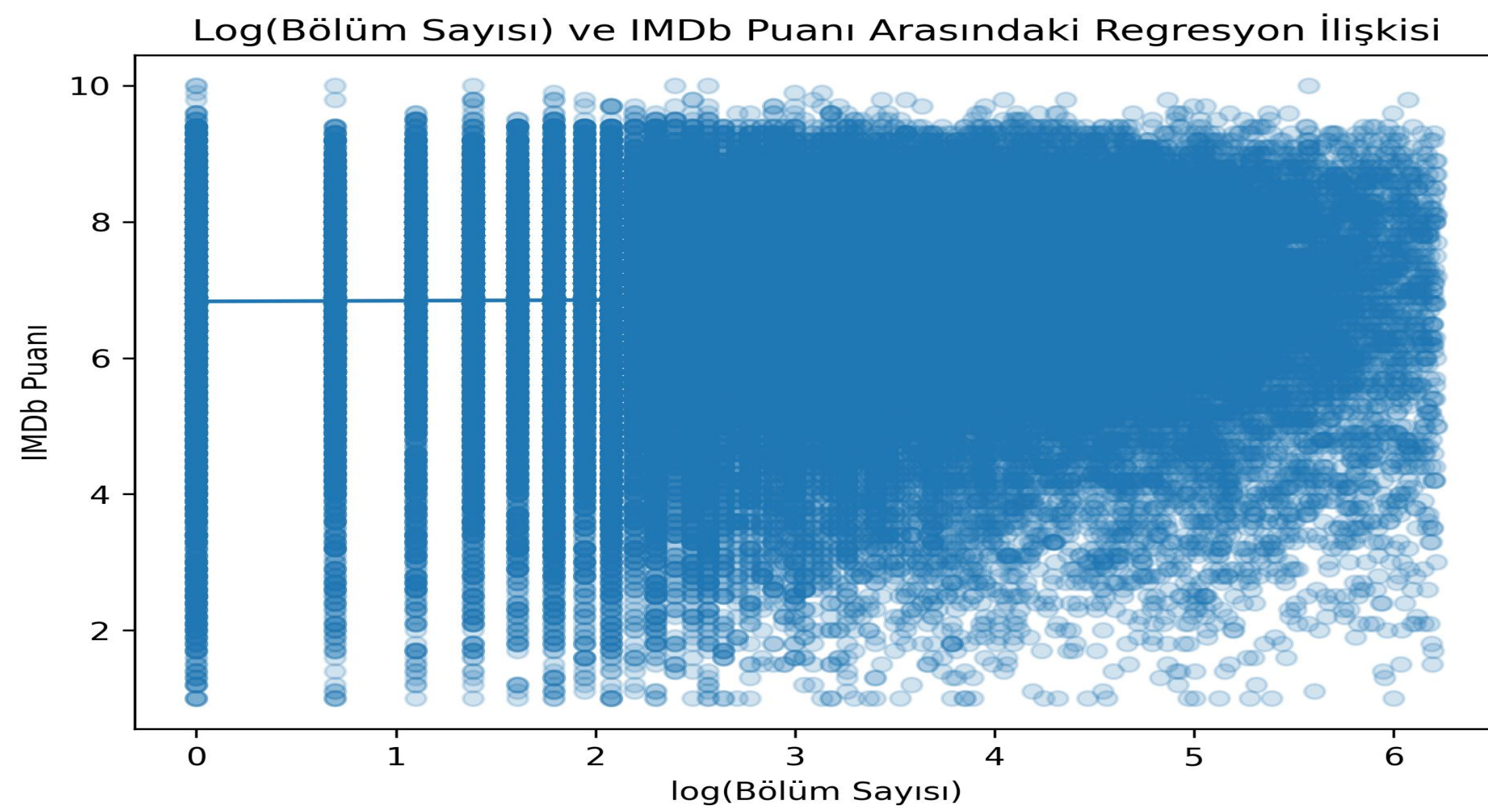
Değişkenler arasındaki ilişkinin yönünü ve şiddetini belirlemek amacıyla Pearson Korelasyon Katsayısı kullanılmıştır. Bölüm sayısının (bağımsız değişken) izleyici puanı (bağımlı değişken) üzerindeki potansiyel etkisini ölçmek, veri setindeki genel eğilimi sayısal olarak ifade etmemize olanak sağlar.



Bölüm sayısı ile izleyici puanı arasındaki doğrusal ilişki korelasyon analizi yardımıyla incelenmiştir. Korelasyon katsayısının sıfıra yakın değerler alması, değişkenler arasında zayıf bir doğrusal ilişki olduğunu göstermektedir. Ancak korelasyon analizi, değişkenler arasındaki nedensel ilişkiyi açıklamak için yeterli olmadığından, sonuçlar regresyon analizi ile desteklenmiştir. Elde edilen sonuçlar da, değişkenler arasındaki ilişkinin zayıf olduğunu göstermektedir ($r = -0.02$).

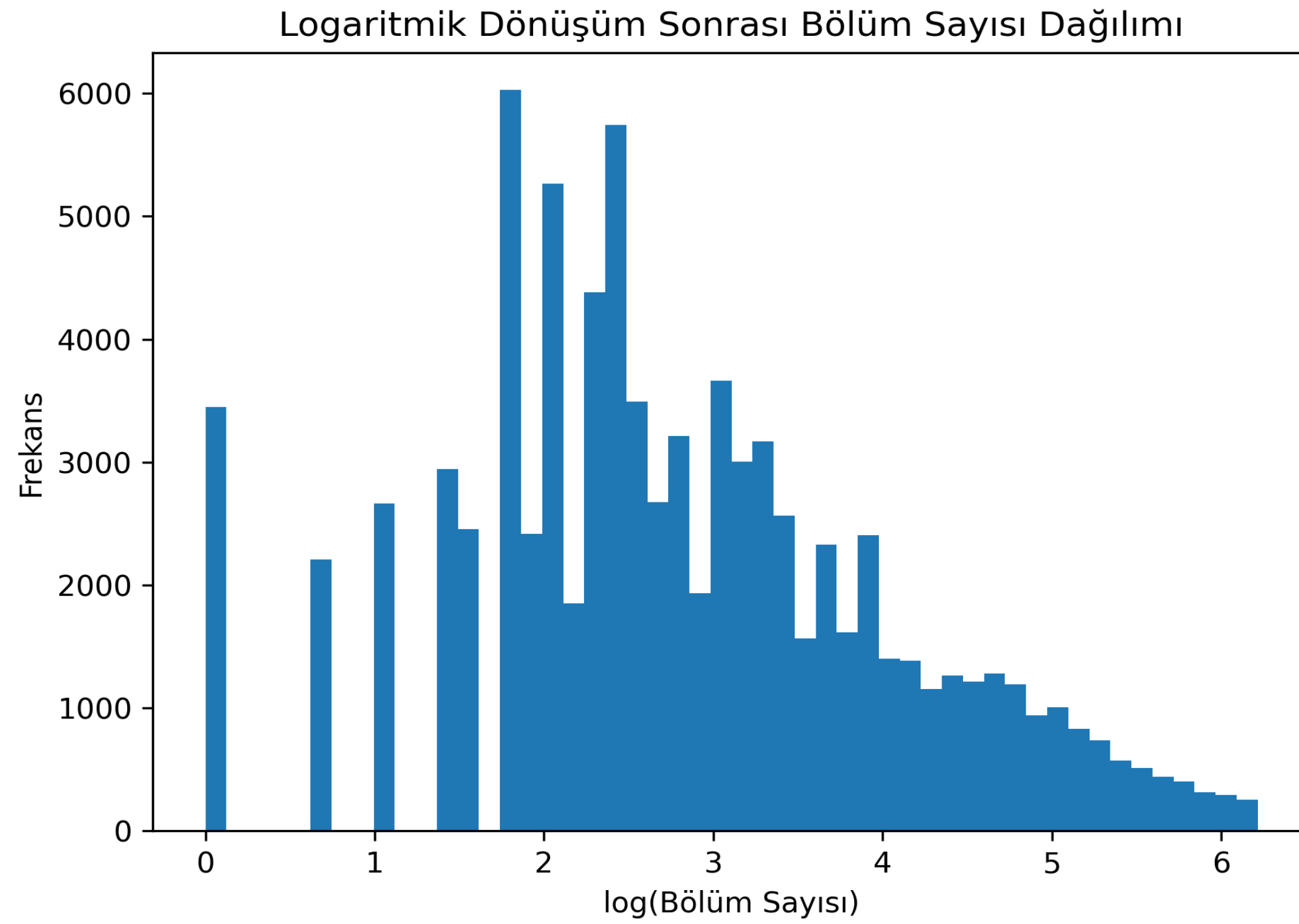
REGRESYON ANALİZİ

Bölüm sayısının izleyici puanı üzerindeki etkisini ölçmek amacıyla doğrusal regresyon modeli kurulmuştur. Regresyon sonuçları, bölüm sayısının izleyici puanları üzerindeki etkisinin istatistiksel olarak anlamlı olmakla birlikte açıklayıcılık düzeyinin oldukça düşük olduğunu göstermektedir.



LOGARİTMİK DÖNÜŞÜM

Bölüm sayısı değişkeninin dağılımı belirgin biçimde çarpık olduğundan, analizlerde logaritmik dönüşüm uygulanmıştır. Bu dönüşüm, uç değerlerin etkisini azaltarak bölüm sayısı ile izleyici puanı arasındaki ilişkinin daha dengeli ve yorumlanabilir biçimde incelenmesini sağlamıştır.

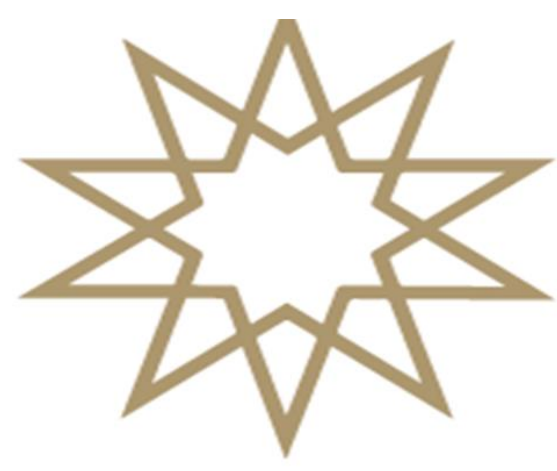


SONUÇ

Bu çalışmada IMDb dizilerinde bölüm sayısı ile izleyici puanı arasındaki ilişki incelenmiştir. Elde edilen bulgular, bölüm sayısının izleyici puanları üzerinde istatistiksel olarak anlamlı olmakla birlikte oldukça sınırlı bir etkiye sahip olduğunu göstermektedir. Korelasyon ve regresyon analizleri, ilişkinin zayıf olduğunu ortaya koymuştur. Bu sonuçlar, izleyici değerlendirmelerinin tek bir nicel değişkenle açıklanamayacak kadar çok boyutlu bir yapıya sahip olduğunu göstermektedir.

KAYNAKÇA

- [1] Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.
- [2] Hill, A. (2005). Reality TV: Audiences and popular factual television. Routledge.
- [3] Cleveland, W. S. (1993). Visualizing data. Hobart Press.



YTU

FEN EDEBİYAT FAKÜLTESİ

İSTATİSTİK BÖLÜMÜ

RFM VE K-MEANS YÖNTEMLERİ İLE MÜŞTERİ SEGMENTASYONU ANALİZİ

Yakup Eşgi - 22023602

Danışman: Dr. Öğr. Üyesi Mehmet Şamil Güneş

ÖZET

Bu çalışmanın amacı, bir e-ticaret firmasına ait sipariş verilerini kullanarak, kural tabanlı RFM (Recency, Frequency, Monetary) analizi ve makine öğrenmesi tabanlı K-Means kümeleme algoritmasını entegre eden hibrit bir segmentasyon modeli oluşturmaktır. Çalışma kapsamında, veri setindeki çarpıklık sorununu gidermek ve algoritma performansını artırmak amacıyla Recency değişkeni için Box-Cox dönüşümü, Frequency ve Monetary değişkenleri için ise Logaritmik (Log1p) dönüşüm uygulanmıştır. Elde edilen ve standartlaştırılan veriler üzerinden optimum küme sayısını belirlemek için Dirsek (Elbow) yöntemi ve Silhouette analizi kullanılmıştır. Analiz sonucunda, müşteriler satın alma alışkanlıklarına ve harcama düzeylerine göre istatistiksel olarak anlamlı 4 farklı stratejik segmente ayrılmıştır. Çalışma sonucunda, RFM ve K-Means yöntemleri ile müşteri segmente edip oluşturulan müşteri grupları için pazarlama stratejileri önerilerek işletmenin kârlılığına ve müşteri sadakatine katkı sağlanması amaçlanmıştır.

Müşteri Segmentasyonu Nedir ?

Müşteri segmentasyonu, bir işletmenin mevcut ve potansiyel müşterilerini; satın alma alışkanlıkları, ihtiyaçları veya davranışsal özellikleri bakımından benzerlik gösteren homojen gruplara ayırma sürecidir. Bu süreç, işletmelerin sınırlı pazarlama kaynaklarını en verimli şekilde kullanmasına, kişiselleştirilmiş hizmet sunarak müşteri memnuniyetini artırmasına ve kârlılığını maksimize etmesine yardımcı olur.

RFM Analizi

RFM analizi, ilk olarak 1994 yılında Arthur Hughes tarafından önerilen ve müşterilerin geçmiş satın alma davranışlarını temel alarak mevcut ve gelecekteki değerlerini ölçümlemeyi amaçlayan, veri madenciliği tabanlı bir segmentasyon tekniğidir. Bu model, müşterileri güncellik (Recency), sıklık (Frequency) ve parasal değer (Monetary) olmak üzere üç temel boyutta değerlendirerek, işletmelerin en kârlı müşteri gruplarını tespit etmesini sağlar.

Recency (Yenilik) Hesabı: Veri setinde bulunan tüm müşteriler içerisinde en güncel tarihli sipariş tarihinin bir gün sonrası Recency değerinin hesaplanması için seçilir. Her bir müşterinin son sipariş tarihi ile hesaplama için seçilen tarih arasındaki fark, gün bazında hesaplanarak bu metrikte depolanır.

Frequency (Sıklık) Hesabı: Müşterilerin işlem kayıtları bulunan tarih aralığında kaç kez sipariş verdiğini belirlemek adına, her bir müşteriye ait sipariş işlemi sayıları toplanarak bu metrik elde edilir.

Monetary (Parasal Değer): Müşterinin işlem kayıtları bulunan tarih aralığında gerçekleştirdiği tüm siparişlerdeki harcanan miktarın toplamı alınarak bu metrik oluşturulur.

K-Means Kümeleme

K-Means, verileri birbirine olan mesafelerine (genellikle Öklid mesafesi) göre "k" sayıda kümeye ayıran gözetimsiz bir algoritmadır. Algoritma, rastgele merkezler seçer, her veriyi en yakın merkeze atar ve merkezleri kümelerin ortalamasına göre güncelleyerek küme içi kareler toplamını (WCSS) minimize edene kadar devam eder.

Veri Seti

Bu çalışmada, e-ticaret müşterilerinin satın alma davranışlarını içeren ve Kaggle platformundan temin edilen bir veri seti kullanılmıştır. Çalışmaya konu olan veri seti, 01.01.2023 – 03.25.2024 tarihleri arasında bulunan 17049 sipariş kaydından ve 18 adet değişkenden oluşmaktadır. Bu değişkenler şunlardır; "Order_id", "Customer_id", "Date", "Age", "Gender", "City", "Product_category", "Unit_price", "Quantity", "Discount_amount", "Total_amount", "Payment_method", "Device_type", "Session_duration_minutes", "Pages_viewed", "İs_returning_customer", "Delivery_time_days", "Customer_rating"

Uygulama

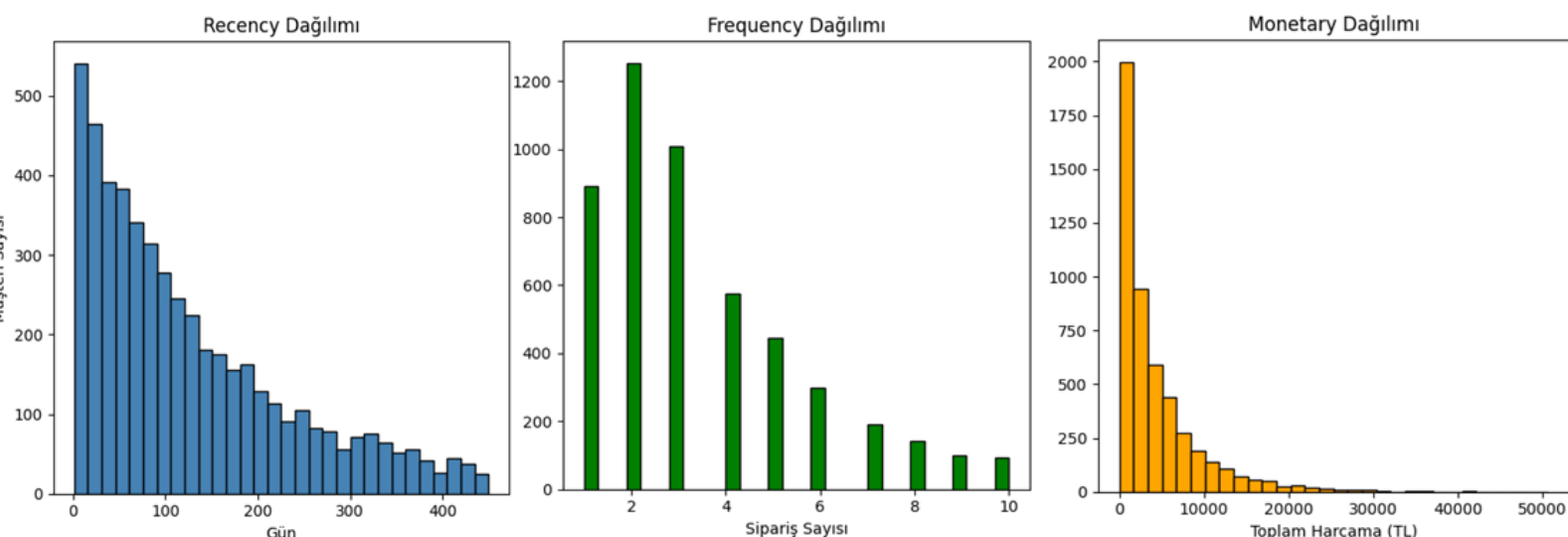
RFM analizi için veri setindeki en güncel sipariş tarihinden bir gün sonrası R parametresinin değerler alabilmesi için seçilir ve metrikler türetilir.

	Customer_ID	Recency	Frequency	Monetary
0	CUST_00001	112	3	2199.630000
1	CUST_00002	284	2	809.900000
2	CUST_00003	83	2	3030.810000

Burada, Recency, son sipariş üzerinden geçen günü;

Frequency, işlem kayıtları bulunan tarih aralığında kaç kez sipariş verildiğini;

Monetary ise aynı tarih aralığındaki harcanan toplam tutarı belirtir.



RFM metriklerinin normal dağılım göstermediği ve her üç değişkenin de sağa çarpık bir yapıya sahip olduğu görülmektedir. Hesaplanan metriklerin, birbirinden farklı ölçeklere (gün, adet, para birimi) sahip oldukları için doğrudan karşılaştırılmazlar. Bu nedenle, metrikler 1 ile 5 arasında değişen

standart skorlara dönüştürülmüştür. Bu işlemde pandas kütüphanesinden “qcut” (quantile cut) yöntemi kullanılarak her bir metrik %20’lik kartiller halinde 5 eşit parçaya bölünmüştür. Böylece her müşteri için “555” (En değerli) ile “111” (En değersiz) arasında değişen bir RFM Skoru oluşturulmuştur.

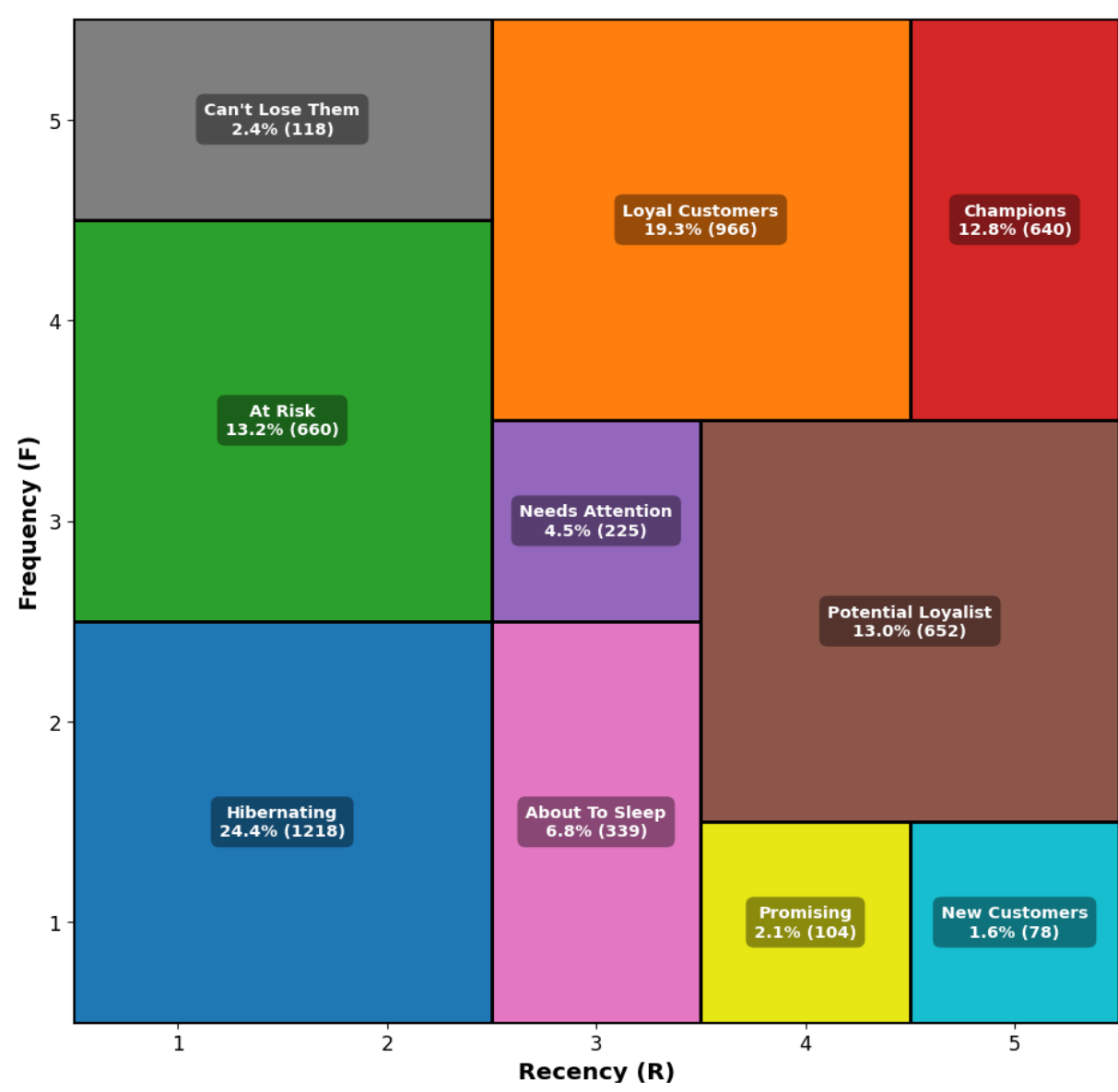
RFM metriklerinin normal dağılım göstermediği ve her üç değişkenin de sağa çarpık bir yapıya sahip olduğu görülmektedir. Hesaplanan metriklerin, birbirinden farklı ölçeklere (gün, adet, para birimi) sahip oldukları için doğrudan karşılaştırılmazlar. Bu nedenle, metrikler 1 ile 5 arasında değişen standart skorlara dönüştürülmüştür. Bu işlemde pandas kütüphanesinden “qcut” (quantile cut) yöntemi kullanılarak her bir metrik %20’lik kartiller halinde 5 eşit parçaya bölünmüştür. Böylece her müşteri için “555” (En değerli) ile “111” (En değersiz) arasında değişen bir RFM Skoru oluşturulmuştur.

	Customer_ID	Recency	Frequency	Monetary	R_Score	F_Score	M_Score	RFM_Score
0	CUST_00001	112	3	2199.630000	3	3	3	333
1	CUST_00002	284	2	809.900000	1	1	2	112
2	CUST_00003	83	2	3030.810000	3	1	3	313

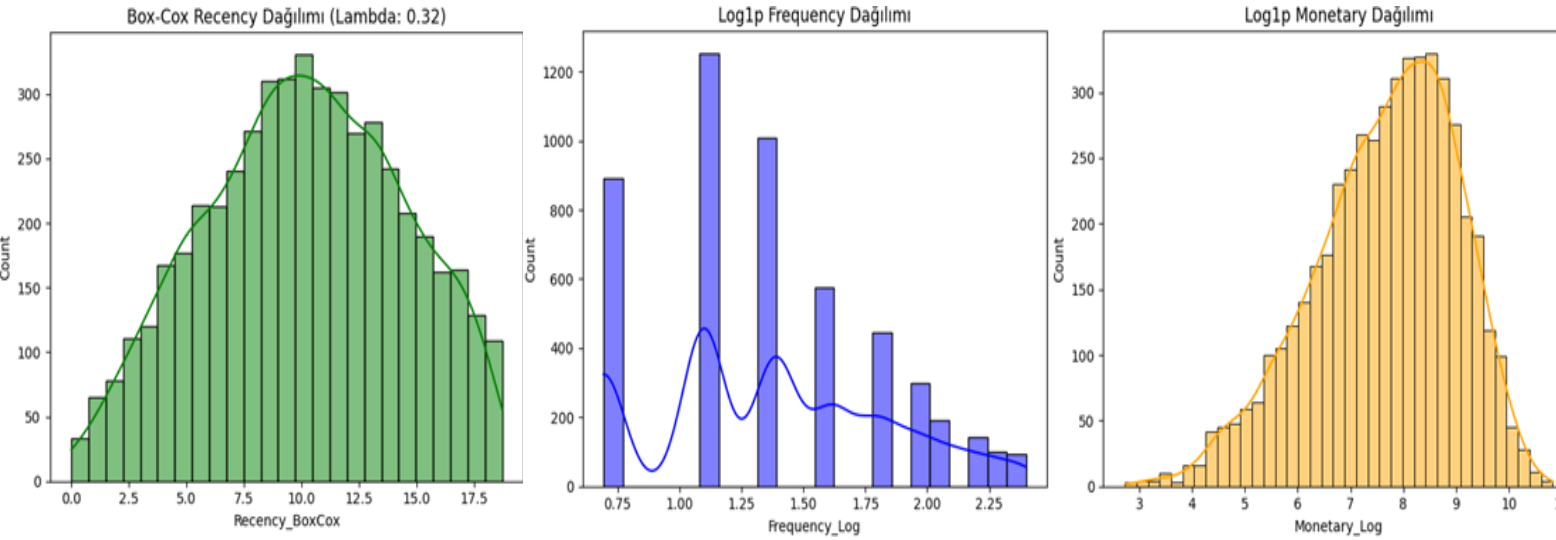
Elde edilen RFM skorları ile müşteriler, aşağıda belirtilen kriterler esas alınarak 10 farklı segmente ayrılmıştır. Bu segmentasyon sürecinde temel belirleyiciler Recency ve Frequency skorları olmuştur. Segmentasyon sonucunda elde edilen tablo ise Ağaç Haritası olarak sunulmuştur.

Segment Adı	Recency Skoru	Frequency Skoru
Champions	5	4 - 5
Loyal Customers	3 - 4	4 - 5
Potential Loyalist	4 - 5	2 - 3
Promising	4	1
Needs Attention	3	3
About To Sleep	3	1 - 2
At Risk	1 - 2	3 - 4
Can't Loose Them	1 - 2	5
Hibernating	1 - 2	1 - 2
New Customers	5	1

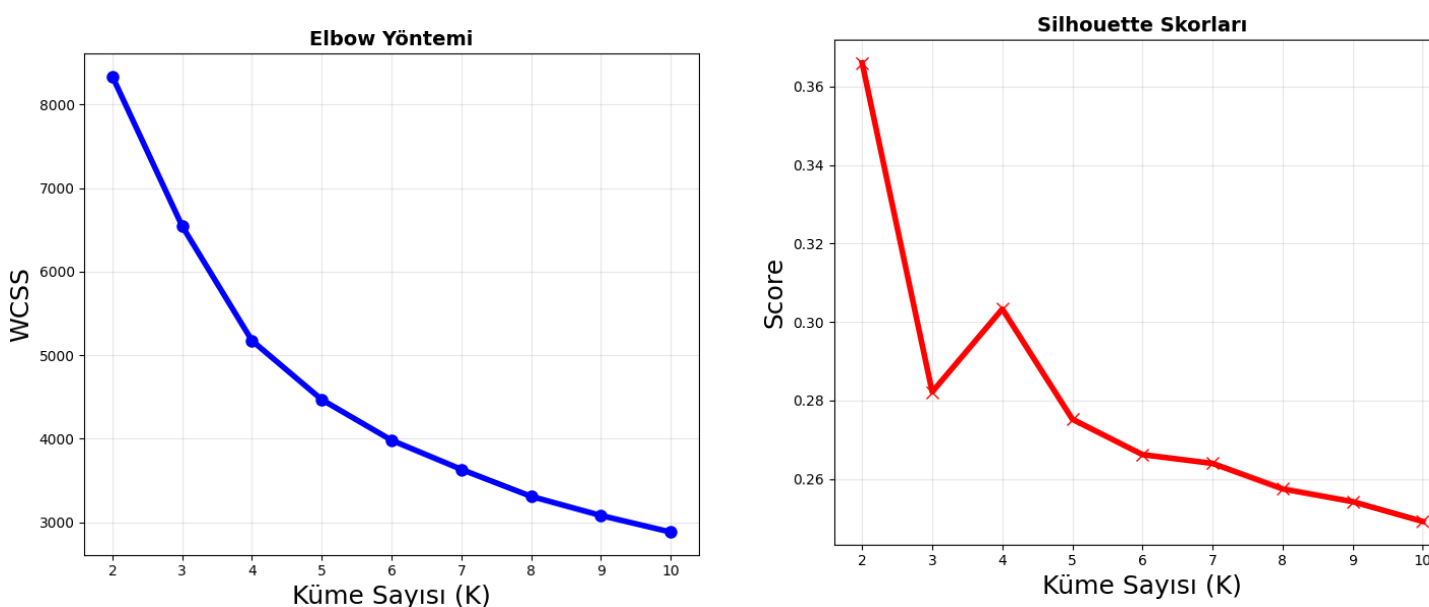
RFM Müşteri Segmentasyon Ağaç Haritası



K-Means algoritması, gözlemler arasındaki uzaklıkları temel alarak çalıştığından, veri setinin dağılımı ve ölçeği sonuçları doğrudan etkilemektedir. Ham RFM metrikleri incelendiğinde, verilerin normal dağılımdan uzak olduğu ve sağa çarpık bir yapı sergilediği görülmüştür. Ayrıca değişkenlerin ölçekleri (gün, adet, TL) birbirinden farklıdır.

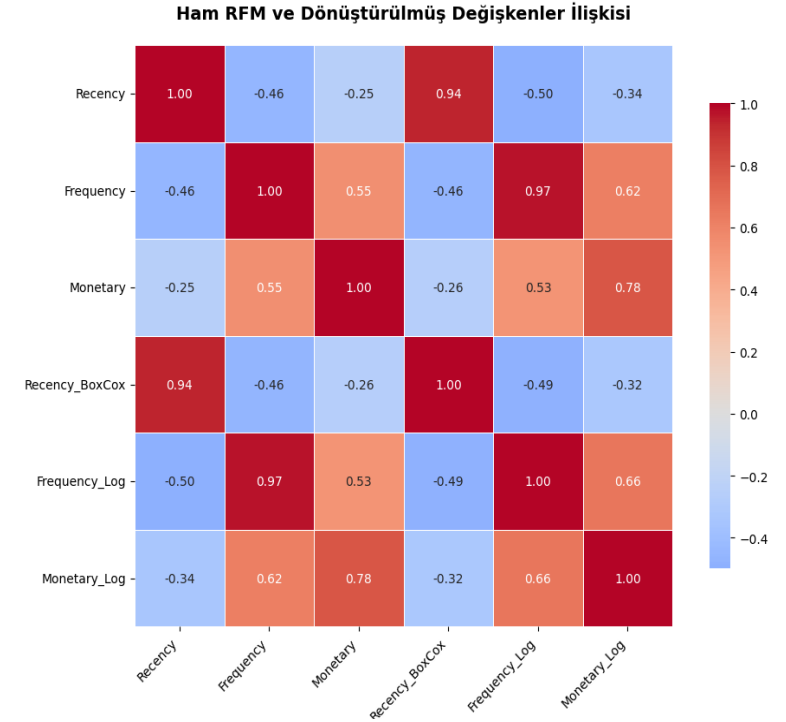


Gözlemlenen güçlü ilişki, dönüştürülmüş değişkenlerin ham metrikleri temsil etme gücünü büyük oranda sürdürdüğünü doğrulamaktadır.

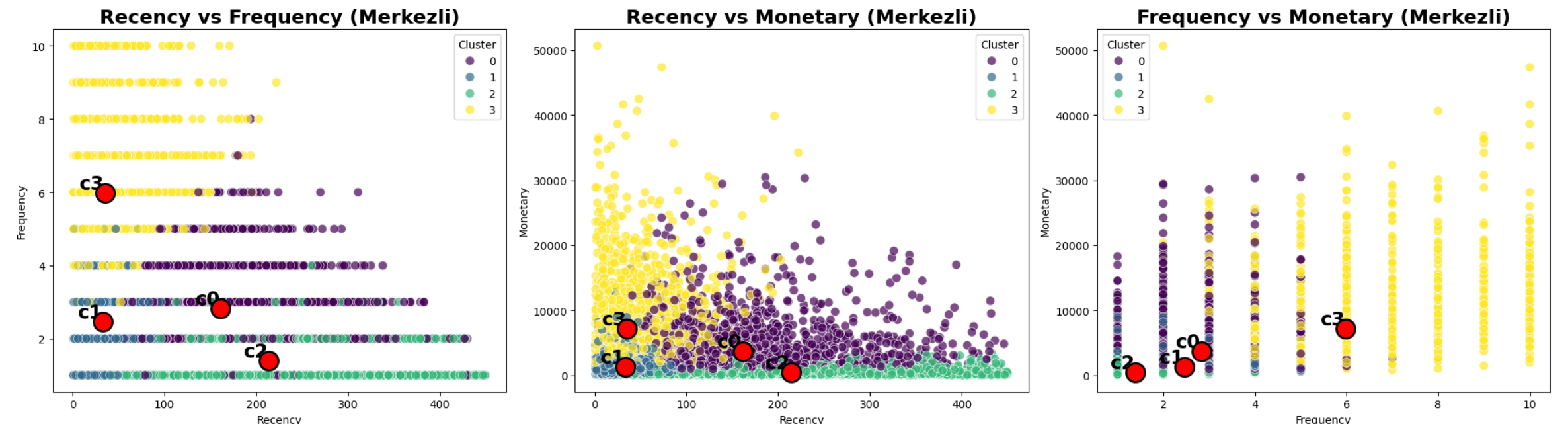


R metriği için Box-Cox, F ve M metrikleri için log1p dönüşümü uygulanmıştır. Dönüşüm sonrası dağılımlar yanda görüldüğü gibidir.

Dönüşüm işlemlerinin veri setinin yapısal bütünlüğünü koruyup korumadığını teyit etmek amacıyla, ham metrikler ile dönüştürülmüş değişkenler arasındaki ilişki katsayıları incelenmiştir.



Çalışmada, en ideal bölümlmeyi sağlayan küme sayısını belirlemek için Elbow yöntemi ve Silhouette analizi birlikte değerlendirilmiş, optimum küme sayısı k=4 olarak belirlenen k=4 değeri ile K-Means algoritması çalıştırılmış ve her bir müşteri ilgili kümeye atanmıştır. Kümelerin gösterimi aşağıdaki gibidir.



DETAYLI KÜME PROFİL ÖZETİ

Küme ID	Kişi Sayısı	Oran (%)	Ort. Recency	Ort. Frequency	Ort. Monetary	Product_Category	Payment_Method	Device_Type
0	1471	29.4%	172.7 Gün	3.0 Adet	4849.15 TL	Electronics (%16), Sports (%15), Home & Garden (%14)	Credit Card (%40), Debit Card (%24), Digital Wallet (%20)	Mobile (%57), Desktop (%33), Tablet (%10)
1	1019	20.4%	40.5 Gün	2.6 Adet	1802.88 TL	Books (%17), Food (%14), Fashion (%14)	Credit Card (%39), Debit Card (%27), Digital Wallet (%19)	Mobile (%55), Desktop (%35), Tablet (%10)
2	1167	23.3%	233.1 Gün	1.5 Adet	589.78 TL	Books (%19), Beauty (%18), Food (%17)	Credit Card (%41), Debit Card (%24), Digital Wallet (%19)	Mobile (%57), Desktop (%33), Tablet (%10)
3	1343	26.9%	45.6 Gün	6.2 Adet	9024.98 TL	Electronics (%14), Sports (%14), Home & Garden (%13)	Credit Card (%40), Debit Card (%26), Digital Wallet (%19)	Mobile (%56), Desktop (%35), Tablet (%10)

Küme - 3 “Şampiyonlar ”, Küme - 2 “Pasif Müşteriler ”, Küme - 1 “Yeni / Potansiyelli Müşteriler ”, Küme - 0 “Riskli Müşteriler ” olarak tanımlanmıştır.

Sonuç

Çalışmanın genel sonucu olarak; firmanın pazarlama kaynaklarını tüm müşterilere eşit dağıtmak yerine, segmentlerin karakteristik özelliklerine göre ayrıştırılması gerektiği ortaya çıkmıştır. Bu bağlamda, şampiyon kitlenin mevcut memnuniyetinin korunmasının yanı sıra yüksek gelirli ancak riskli olan grubun geri kazanılmasına ve yeni gelen düşük bütçeli grubun sepet tutarı ortalamasının artırılmasına odaklanılması, işletmenin uzun vadeli kârlılığı ve müşteri yaşam boyu değerinin maksimize edilmesi açısından optimum strateji olacaktır.

KAYNAKÇA

- [1] Başkol, M. (2020). "RFM ve uyum analizi kullanılarak müşteri segmentasyonunun belirlenmesi". *Business & Management Studies: An International Journal*, 8(4), ss.902-928.
- [2] Metin, S. (2025). "E-Ticarette kullanıcılarının davranışsal yaklaşımlarının makine öğrenmesi yöntemleri ile sınıflandırılması ve müşteri segmentasyonu". *Hitit Sosyal Bilimler Dergisi*, 18(2), ss.314-329.
- [3] Aslantaş, G., Gençgöl, M. v.d. (2023). "Customer Segmentation Using K-Means Clustering Algorithm and RFM Model". *DEUFMD*, 25(74), ss.491-503.
- [4] UmutUygur. (2025). “E-Commerce Customer Behavior & Sales Analysis -TR” [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/13664353>