



**softITO 4. DÖNEM 2. GRUP
VERİ ANALİTİĞİ
BİTİRME PROJESİ RAPORU**

**BORSA EMİR DEFTERİNDE SPOOFING VE LAYERING
MANİPÜLASYON ÖRÜNTÜLERİNİN ETİKETSİZ
MAKİNE ÖĞRENMESİ İLE TESPİTİ**

Hazırlayan
Abdul Samed Yorulmaz

İÇİNDEKİLER

1. GİRİŞ

- 1.1. Projenin Amacı
- 1.2. Projenin Kapsamı ve Kısıtları

2. MATERYAL VE YÖNTEM

- 2.1. Veri Seti
- 2.2. Limit Emir Defteri Değişkenleri
- 2.3. Order Lifecycle
- 2.4. Emir Defteri Bağlam Özellikleri
- 2.5. Davranışsal Aday Kriterleri
- 2.6. DBSCAN ile Episode Oluşturma
- 2.7. Control Episode Havuzu
- 2.8. Episode-Level Feature Engineering
- 2.9. Isolation Forest
- 2.10. Şüpheli Olayların Önceliklendirilmesi
- 2.11. Fiyat Etkisi ve Reversal
- 2.12. Matched Control ve Event Study
- 2.13. İstatistiksel Değerlendirme

3. UYGULAMA VE BULGULAR

4. SONUÇ VE DEĞERLENDİRME

- 4.1. Gelecek Çalışmalar

5. KAYNAKÇA

1. GİRİŞ

Finansal piyasalarda alım ve satım emirlerinin önemli bir bölümü limit emir defteri (Limit Order Book, LOB) üzerinde gerçekleşmektedir. Limit emir defteri, farklı fiyat seviyelerinde bekleyen alım ve satım emirlerini içererek piyasanın anlık arz ve talep yapısını yansıtır. Elektronik işlem sistemlerinin gelişmesiyle emir gönderme, değiştirme ve iptal işlemleri milisaniye düzeyinde gerçekleşebilmekte; bu durum piyasa mikro yapısının ayrıntılı incelenmesini mümkün kılarken manipülatif emir davranışlarının tespitini de zorlaştırmaktadır.

Spoofing, genel olarak gerçekleşme amacı düşük veya bulunmayan emirlerin piyasaya verilerek arz-talep görünümünün etkilenmesi ve bu emirlerin gerçekleşmeden önce iptal edilmesiyle ilişkilendirilen bir davranıştır. Layering ise benzer davranışın aynı tarafta birden fazla fiyat seviyesine yayılan emir katmanları biçiminde görülmesidir (U.S. Securities and Exchange Commission, 2026; Commodity Futures Trading Commission, t.y.). Her büyük veya hızlı iptal edilen emir manipülasyon anlamına gelmediğinden, tespit sürecinde emir büyüklüğü, yaşam süresi, iptal ve gerçekleşme oranları, emir defterindeki konum ve olay çevresindeki piyasa davranışı birlikte değerlendirilmelidir.

Bu çalışmada Apple Inc. (AAPL) hissesine ait 10 seviyeli LOBSTER limit emir defteri verisi kullanılmıştır. Veri kümesinde gerçek spoofing/layering etiketleri bulunmadığı için problem supervised classification olarak değil, davranışsal örüntü analizi ve etiketsiz anomali tespiti problemi olarak ele alınmıştır.

1.1. Projenin Amacı

Çalışmanın temel amacı, AAPL hissesine ait yüksek zaman çözünürlüklü emir defteri verisinde, gerçek manipülasyon etiketi kullanılmadan spoofing ve layering karakteristiği taşıyan şüpheli olayları belirlemek ve bu olayları inceleme önceliğine göre sıralamaktır. Araştırma; davranışsal kriterlerle aday emirlerin belirlenmesi, ilişkili emirlerin episode düzeyinde gruplanması, candidate ve normal piyasa episode'larının ortak referans dağılımında anomaliliklerinin ölçülmesi ve olay çevresindeki fiyat davranışının matched-control yaklaşımıyla incelenmesi üzerine kurulmuştur.

1.2. Projenin Kapsamı ve Kısıtları

Çalışma tek bir AAPL işlem gününün 10 seviyeli LOBSTER verisiyle sınırlıdır. Ground-truth manipülasyon etiketi ve trader identity bilgisi bulunmamaktadır. Bu nedenle sonuçlar kesin veya hukuki manipülasyon tespiti olarak yorumlanamaz. DBSCAN parametreleri ve behavioral threshold seçimleri episode havuzunu etkileyebilir. Isolation Forest yalnızca istatistiksel sıra dışılığı ölçer; piyasa katılımcısının niyetini belirleyemez. Event-study sonuçları da pre-event farklılıklar nedeniyle güçlü nedensellik kanıtı olarak değerlendirilmemiştir.

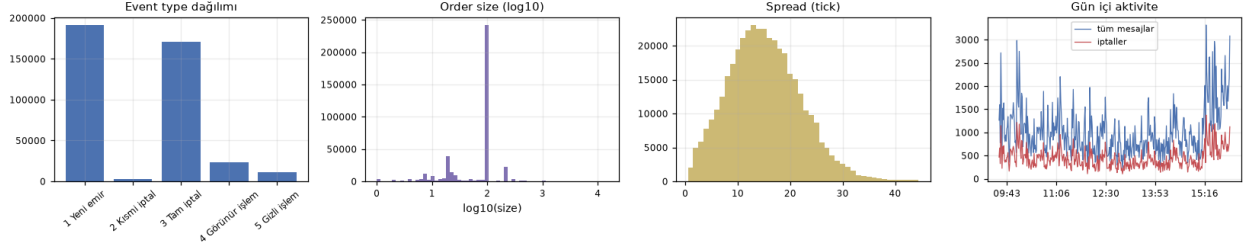
2. MATERYAL VE YÖNTEM

2.1. Veri Seti

LOBSTER, NASDAQ TotalView-ITCH mesajlarından limit emir defterini yeniden oluşturan bir veri sistemidir (LOBSTER, t.y.). Message verisi zaman, olay türü, emir kimliği, hacim, fiyat ve yön bilgilerini; order book verisi ise farklı seviyelerdeki alış-satış fiyatları ve hacimleri içerir. Çalışmada düzenli piyasa saatleri (09:30-16:00) ve ilk 10 emir defteri seviyesi kullanılmıştır.

Tablo 1. LOBSTER olay türleri.

Event Type	Açıklama
1	Yeni limit emri
2	Kısmi iptal
3	Emir silme
4	Görünür limit emrin gerçekleşmesi
5	Gizli emrin gerçekleşmesi



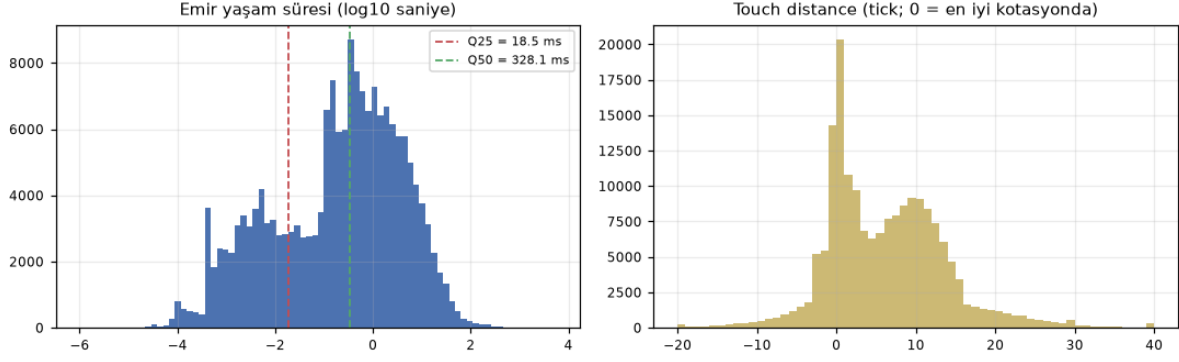
Şekil 1. AAPL LOBSTER verisinin temel keşifsel veri analizi görünümleri.

2.2. Limit Emir Defteri Değişkenleri

En iyi alış ve satış fiyatlarından mid-price ve spread; ilk beş seviyedeki alış-satış hacimlerinden ise order-book imbalance ve görünür derinlik değişkenleri türetilmiştir. Mid-price $(\text{Bid} + \text{Ask})/2$, spread ise $\text{Ask} - \text{Bid}$ olarak hesaplanmıştır.

2.3. Order Lifecycle

Her order_id için yeni emir kaydı başlangıç noktası alınmış; iptal ve gerçekleşme olayları kümülatif olarak takip edilmiştir. Kalan hacim, $\text{InitialVolume} - \text{CancelledVolume} - \text{ExecutedVolume}$ olarak hesaplanmış ve emir ancak kalan hacim sıfıra ulaştığında kapanmış kabul edilmiştir. Böylece kısmi iptal ve kısmi gerçekleştirmeler yanlış biçimde terminal olay olarak değerlendirilmemiştir.



Şekil 2. Kapanmış emirlerde yaşam süresi ve touch-distance dağılımları.

2.4. Emir Defteri Bağlam Özellikleri

Emrin kendi tarafındaki en iyi kotasyona uzaklığı tick cinsinden touch distance ile, emrin görünür piyasa derinliğine göre göreceli büyüklüğü ise size_vs_depth değişkeniyle temsil edilmiştir. Bu yaklaşım mutlak olarak büyük olmayan fakat mevcut likiditeye göre sıra dışı büyüklükte olan emirlerin de değerlendirilmesini sağlar.

2.5. Davranışsal Aday Kriterleri

Gerçek etiket bulunmadığından ilk aşamada behavioral candidate yaklaşımı kullanılmıştır. Temel göstergeler kısa yaşam süresi, büyük veya derinliğe göre büyük emir, yüksek iptal oranı, düşük gerçekleşme oranı ve layering için kısa zaman penceresinde birden fazla fiyat seviyesine yayılan emir davranışlarıdır. Eşikler mümkün olduğunca veri dağılımının quantile değerlerinden türetilmiştir.

2.6. DBSCAN ile Episode Oluşturma

Behavioral candidate emirler BUY ve SELL taraflarında ayrı ayrı, zaman ve fiyat yakınlığına göre DBSCAN ile gruplanmıştır (Ester et al., 1996). DBSCAN burada manipülasyon veya anomali tespit eden model değildir; ilişkili emirleri tek bir episode altında toplamak için kullanılmıştır. Noise noktaları anomali kabul edilmemiş, tek-emirlik episode olarak korunmuştur. Final uygulamada eps=0,3 ve min_samples=2 kullanılmış; zaman 1 saniye, fiyat 10 tick ölçeğine göre normalize edilmiştir.

2.7. Control Episode Havuzu

Selection bias riskini azaltmak için Isolation Forest yalnızca önceden şüpheli seçilmiş episode'larda çalıştırılmamıştır. Candidate olmayan emirler de aynı DBSCAN tanımıyla control episode'lara dönüştürülmüş ve model candidate + control ortak havuzunda fit edilmiştir.

2.8. Episode-Level Feature Engineering

Her episode; emir sayısı, fiyat seviyesi sayısı, toplam hacim, episode süresi, emir yaşam süreleri, iptal/gerçekleşme oranları, touch distance, spread, imbalance, piyasa aktivitesi ve benzeri özelliklerle sabit uzunlukta bir feature vektörüne dönüştürülmüştür. Candidate/control veya spoofing/layering bilgileri feature olarak verilmemiştir. Post-event price impact ve reversal değişkenleri de future leakage yaratmamak için modele dahil edilmemiştir.

2.9. Isolation Forest ile Anomali Tespiti

Isolation Forest, gözlemleri rastgele bölmelerle izole ederek sıra dışı örneklerin daha kısa izolasyon yollarına sahip olması ilkesine dayanan etiketsiz bir anomali tespit yöntemidir (Liu et al., 2008). Model candidate ve control episode ortak havuzunda çalıştırılmış, her episode için sürekli anomaly score üretilmiştir.

contamination=0,05 yalnızca operasyonel anomali bayrağı eşiği olarak kullanılmış; gerçek manipülasyon oranı olarak yorumlanmamıştır. Tekrar üretilebilirlik için random_state=42 sabitlenmiştir.

2.10. Şüpheli Olayların Önceliklendirilmesi

Final değerlendirme üç bileşenden oluşmaktadır: Behavioral Evidence, Anomaly Evidence ve Price-Pattern Signal. Üçüncü bileşen nedensellik veya manipülasyon kanıtı değil, olay çevresindeki fiyat davranışının hipotezle uyumunu gösteren destekleyici sinyaldir. Candidate episode'lar High-Priority Spoofing, High-Priority Layering, Moderate ve Weak Candidate kategorilerine ayrılmıştır.

2.11. Fiyat Etkisi ve Reversal Analizi

Modelleme tamamlandıktan sonra olay çevresindeki mid-price kullanılarak kısa vadeli signed price impact ve reversal ölçümleri hesaplanmıştır. Bu değişkenler Isolation Forest eğitiminde kullanılmamış, yalnızca post-hoc piyasa davranışı analizi amacıyla değerlendirilmiştir.

2.12. Matched Control ve Event Study

Candidate episode'lar benzer piyasa koşullarındaki normal episode'larla eşleştirilmiş; event-study yaklaşımı finans literatüründeki olay çevresi fiyat davranışı çerçevesine uygun biçimde uygulanmıştır (MacKinlay, 1997). olay öncesi, olay sırası ve olay sonrası fiyat yolları karşılaştırılmıştır. Böylece yalnızca candidate olaydan sonra fiyatın hareket edip etmediği değil, benzer normal bir olaydan farklı hareket edip etmediği incelenmiştir.

2.13. İstatistiksel Değerlendirme

Eşleştirilmiş candidate-control çiftleri Wilcoxon Signed-Rank Test ile karşılaştırılmış (Wilcoxon, 1945) ve farkın büyüklüğü matched rank-biserial correlation ile raporlanmıştır. Ayrıca olay başlamadan önce grupların zaten farklı olup olmadığını kontrol etmek için pre-event placebo analizi yapılmıştır.

3. UYGULAMA VE BULGULAR

3.1. Veri ve Lifecycle Bulguları

Tablo 2. Veri ve order lifecycle büyüklükleri.

Ölçü	Değer
Ham piyasa olayı	400.391
Order lifecycle	191.015
Analize alınan kapanmış emir	180.928
Gün sonunda açık kalan emir	10.087

Lifecycle analizi 164.127 emrin tamamen iptal, 15.946 emrin tamamen gerçekleşmiş, 855 emrin ise kısmen gerçekleşip kapanmış olduğunu göstermiştir. Yüksek iptal davranışının normal piyasa emirlerinde de yaygın olması, tek başına iptal oranının manipülasyon göstergesi olamayacağını desteklemektedir.

3.2. Behavioral Candidate Sonuçları

Tablo 3. Behavioral candidate emir sayıları.

Candidate türü	Emir sayısı
Spoofing candidate order	7.726
Layering candidate order	2.702
Benzersiz toplam candidate	9.108

Candidate emirler kapanmış emirlerin yaklaşık %5,03'ünü oluşturmaktadır. Candidate emirlerin medyan büyüklüğü 200, normal emirlerin 100 hissedir. Candidate yaşam süresi medyanı yaklaşık 0,0023 saniye iken normal emirlerde yaklaşık 0,3621 saniyedir.

3.3. DBSCAN Episode Sonuçları

9.108 behavioral candidate emir DBSCAN sonrasında 3.231 candidate episode'a dönüştürülmüştür. Bunların 1.014'ü çoklu emir kümesi, 2.217'si ise tek-emirlik episode olarak korunan noise noktalarıdır.

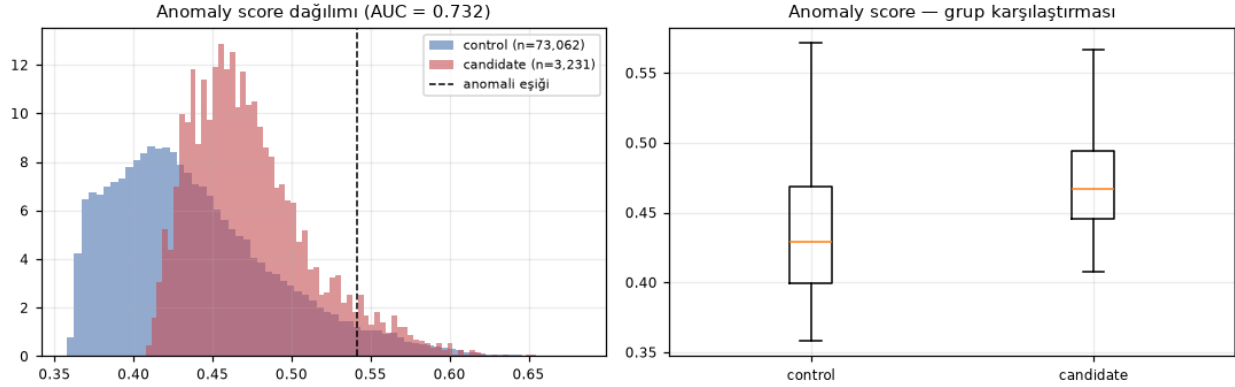
3.4. Candidate-Control Ortak Havuzu

Tablo 4. Candidate-control episode havuzu.

Grup	Episode sayısı
Candidate	3.231
Control	73.062
Toplam	76.293

3.5. Isolation Forest Bulguları

Isolation Forest toplam 76.293 episode üzerinde çalıştırılmış ve 3.815 episode anomali bayrağı almıştır. Candidate episode'ların ortalama anomaly score'u 0,4742, control episode'ların ortalaması 0,4388 olarak bulunmuştur. Candidate-control anomaly-score dağılımlarının sıra tabanlı ayırma ölçüsü AUC=0,7322'dir. Bu değer accuracy değildir; ground-truth etiket bulunmadığı için klasik supervised performans metriği olarak yorumlanmamıştır.



Şekil 3. Candidate ve control episode anomaly-score dağılımlarının karşılaştırılması.

3.6. Nihai Candidate Önceliklendirmesi

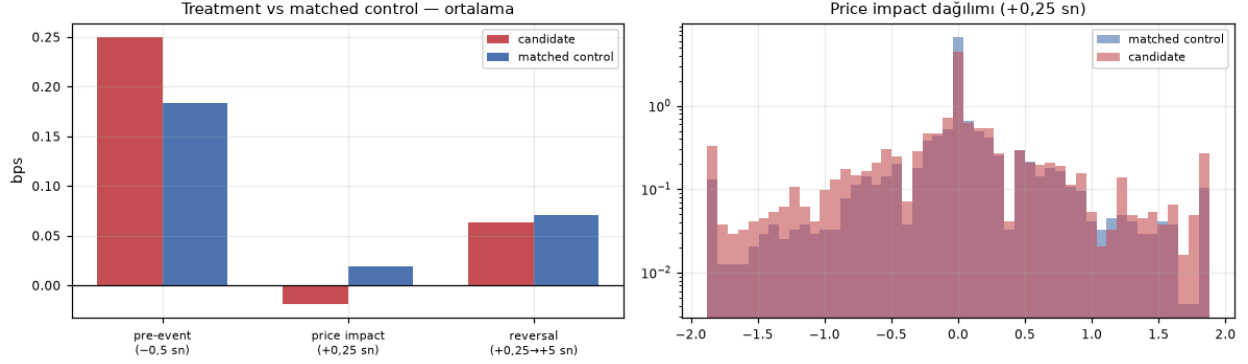
Tablo 5. Nihai candidate önceliklendirmesi.

Final değerlendirme	Episode sayısı
High-Priority Spoofing Candidate	57
High-Priority Layering Candidate	44
Moderate Candidate	1.402
Weak Candidate	1.728
Toplam	3.231

Toplam 101 episode behavioral özellikler, yüksek anomalilik ve destekleyici fiyat örüntüsü birlikte değerlendirildiğinde High-Priority Candidate olarak belirlenmiştir. Bu sayı kesin manipülasyon vakası sayısı değil, daha ileri inceleme için önceliklendirilmiş olay sayısıdır.

3.7. Matched Control ve Event Study

3.231 candidate episode'un 3.153'ü uygun bir control episode ile eşleştirilmiş ve yaklaşık %97,6 eşleşme oranı elde edilmiştir. Price impact ve reversal sonuçları aşağıdaki şekilde karşılaştırılmıştır.



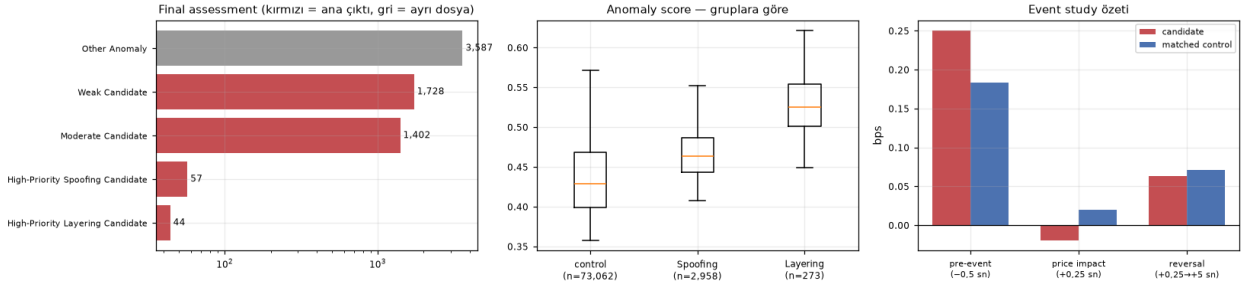
Şekil 4. Candidate ve matched-control gruplarında event-study fiyat ölçümleri.

3.8. İstatistiksel Sonuçlar ve Nedensellik

Tablo 6. Eşleştirilmiş istatistiksel test sonuçları.

Analiz	Wilcoxon p	Rank-biserial	Yorum
Pre-event placebo	0,000518	+0,0783	Fark var; etki çok küçük
Price impact	0,00413	-0,0682	Fark var; etki çok küçük ve ters yönde
Reversal	0,673	-0,0091	Anlamlı fark yok

Price impact farkı istatistiksel olarak anlamlı olmakla birlikte etki büyüklüğü çok küçüktür ve işaret spoofinghipotezinin beklediği yönün tersidir. Reversal açısından anlamlı fark bulunmamıştır. Daha önemlisi, pre-event placebo testi candidate ve control gruplarının olay başlamadan önce de farklı olduğunu göstermiştir. Bu nedenle event sonrasında gözlenen fiyat farkının candidate olaya nedensel olarak atfedilmesi mümkün değildir.



Şekil 5. Final değerlendirme, anomaly-score grupları ve event-study sonuçlarının özeti.

4. SONUÇ VE DEĞERLENDİRME

Bu çalışmada AAPL LOBSTER emir defteri verisinde, gerçek spoofing/layering etiketi olmadan şüpheli emir davranışlarını belirlemek için çok aşamalı bir yaklaşım uygulanmıştır. Behavioral kriterler sonucunda 9.108 benzersiz candidate emir, DBSCAN sonrasında 3.231 candidate episode elde edilmiştir. Candidate olmayan piyasa davranışlarından aynı episode tanımıyla 73.062 control episode oluşturulmuş ve Isolation Forest 76.293 episode'luk ortak referans dağılımında çalıştırılmıştır.

Candidate episode'ların ortalama anomaly score'unun control grubundan yüksek olması ve $AUC=0,7322$ düzeyindeki dağılım ayrışması, behavioral kriterlerle seçilen olayların candidate bilgisini görmeyen etiketsiz anomaly modeli tarafından da daha sıra dışı bulunduğunu göstermektedir. Çalışmanın en güçlü bulgusu behavioral evidence ile anomaly evidence'ın birbirini bağımsız olarak desteklemesidir.

Final önceliklendirmede 57 High-Priority Spoofing ve 44 High-Priority Layering olmak üzere 101 yüksek öncelikli inceleme adayı belirlenmiştir. Bununla birlikte matched-control event study, bu olayların fiyat üzerinde sistematik ve kendi lehlerine nedensel bir etki oluşturduğunu desteklememiştir. Price impact farkının etki büyüklüğü çok küçük, reversal farkı anlamsız ve pre-event placebo farkı anlamlıdır. Bu nedenle çalışma sıra dışı ve manipülasyon karakteristiği taşıyan emir davranışlarını tespit edebilmekte; ancak bunların gerçek manipülasyon olduğunu veya fiyatı nedensel olarak etkilediğini kanıtlamamaktadır.

Sonuç olarak geliştirilen yapı bir otomatik manipülasyon karar mekanizması değil, piyasa gözetiminde daha ileri insan incelemesine yönlendirilecek olayları önceliklendiren bir karar destek yaklaşımıdır. Ground-truth etiket bulunmaması nedeniyle accuracy, precision, recall ve F1 gibi supervised metrikler hesaplanmamıştır. DBSCAN ve behavioral threshold parametrelerinin sonuçlara etkisi, trader identity bilgisinin bulunmaması ve tek hisse/tek gün kapsamı çalışmanın temel sınırlılıklarıdır.

4.1. Gelecek Çalışmalar

Gelecek çalışmalarda yöntemin birden fazla işlem günü ve farklı hisselerde test edilmesi, farklı volatiliteler ve likidite rejimlerinde eşiklerin kararlılığının incelenmesi ve mümkünse trader identity veya uzman doğrulamalı manipülasyon etiketleri içeren veri kaynaklarıyla değerlendirme yapılması önerilmektedir. Gerçek etiket elde edilmesi durumunda mevcut önceliklendirme yaklaşımı precision, recall ve F1 gibi metriklerle doğrudan doğrulanabilir. Daha güçlü nedensel çıkarım yöntemleri ise ancak gerekli varsayımlar veri üzerinde savunulabildiği ölçüde eklenmelidir.

5. KAYNAKÇA

1. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of KDD-96, 226-231.
2. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. 2008 Eighth IEEE International Conference on Data Mining, 413-422.
3. Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83.
4. MacKinlay, A. C. (1997). Event studies in economics and finance. Journal of Economic Literature, 35(1), 13-39.
5. LOBSTER. Data Structure and Output Documentation. Limit Order Book System - The Efficient Reconstructor.
6. LOBSTER. How Does LOBSTER Work? Limit Order Book System - The Efficient Reconstructor.
7. U.S. Securities and Exchange Commission. (2026). Complaint and market surveillance materials concerning spoofing and layering.
8. Commodity Futures Trading Commission. Disruptive Trading Practices / Spoofing Rules.