

Search

Get app

Write



Home

Library

Profile

Stories

Stats

Following

Medium Staff

Find writers and publications to follow.

[See suggestions](#)

Welcome Offer

Access to everything. Now 30% off. [Upgrade now](#)

Machine Learning ✓

Data Science ✓

Anomaly Detection +

Algorithmic Trading +

Financial Markets +

Etiketsiz Emir Defteri Verisinde Spoofing ve Layering Tespiti



Samet Yorulmaz · 6 min read · Just now



AAPL LOBSTER verisi üzerinde davranışsal örüntüler, DBSCAN ve Isolation Forest ile bir piyasa gözetim yaklaşımı

Elektronik piyasalarda manipülatif davranışları tespit etmek, yalnızca fiyat grafiğinde sıra dışı bir hareket aramaktan ibaret değil. Özellikle spoofing ve layering gibi emir defteri manipülasyonlarında asıl iz, emirlerin nasıl yerleştirildiğinde, ne kadar süre yaşadığında, ne ölçüde gerçekleştiğinde ve ne kadar hızlı geri çekildiğinde saklı.

Bu çalışmada AAPL hissesine ait milisaniye çözünürlüklü LOBSTER limit emir defteri verisini

kullanarak, gerçek spoofing/layering etiketleri olmadan şüpheli emir davranışlarını belirleyen ve bu olayları genel piyasa davranışına göre anomalilik düzeyiyle önceliklendiren bir pipeline geliştirdim.

Amaç, “manipülasyonu otomatik olarak kanıtlayan” bir model kurmak değil; binlerce piyasa olayı arasından insan incelemesine değer olayları sistematik biçimde daraltmak.

1. Emir defterinde manipülasyon sinyali neye benzer?

Spoofing, piyasanın bir tarafında gerçekleşme niyeti düşük veya bulunmayan emirlerin görünür arz-talep dengesini etkileyebilecek biçimde yerleştirilmesi ve gerçekleşmeden önce geri çekilmesiyle ilişkilendirilen bir davranıştır. Layering ise bu yapının birden fazla fiyat seviyesine yayılan emir katmanları biçiminde görülmesidir.

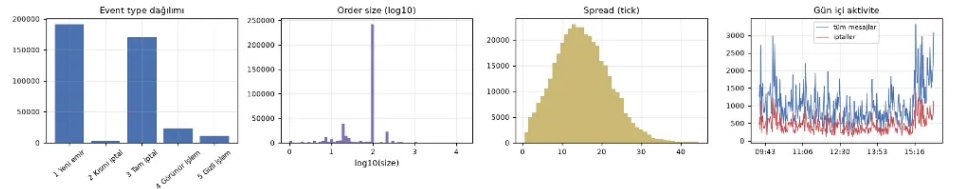
Ancak bu tanımlar gerçek veri üzerinde tek başına yeterli değildir. Büyük emirler meşru olabilir. Hızlı iptal, piyasa yapıcılığında olağandır. Birden fazla fiyat seviyesinde emir bulunması da tek başına manipülasyon anlamına gelmez. Bu nedenle tespit yaklaşımının tek bir kurala değil, birden fazla davranışsal sinyale dayanması gerekir.

Çalışmada bu nedenle üç ayrı kanıt katmanını birbirinden ayırdım: davranışsal yapı, istatistiksel anomalilik ve olay çevresindeki fiyat örüntüsü. Her katman farklı bir soruya cevap veriyor ve hiçbirisi tek

başına “manipülasyon kanıtı” olarak kullanılmıyor.

2. Veri: AAPL LOBSTER limit emir defteri

Analiz, AAPL hissesinin 10 seviyeli LOBSTER limit emir defteri verisi üzerinde yürütüldü. Ham veri 400.391 piyasa olayından oluşuyor. Her olayda zaman, emir türü, order_id, hacim, fiyat ve yön bilgisine ek olarak bid/ask seviyeleri ve bu seviyelerdeki hacimler bulunuyor.



Şekil 1 — LOBSTER verisinin olay türleri, emir büyüklüğü, spread ve gün içi aktivite dağılımları.

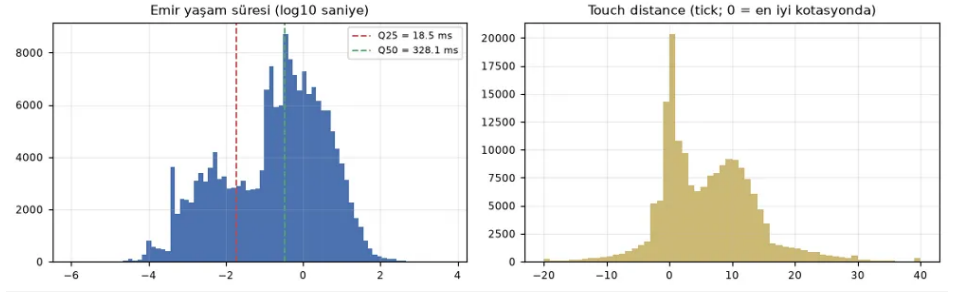
İlk keşifsel analiz önemli bir noktayı ortaya koydu: emir iptali bu piyasada zaten çok yaygın. Bu nedenle yalnızca “yüksek cancellation ratio” gibi tek değişkenli bir kural, normal piyasa davranışı ile manipülasyon karakteristiğini ayırmak için yeterli değil.

3. Analiz birimini doğru kurmak: order lifecycle

Ham event satırlarını doğrudan modele vermek yerine önce her emrin yaşam döngüsünü yeniden oluşturdum. Çünkü bir emrin davranışı, yalnızca gönderildiği anda değil; daha sonra ne kadarının iptal edildiği, ne kadarının gerçekleştiği ve ne kadar süre piyasada kaldığıyla anlam kazanıyor.

Her order_id için başlangıç hacmi, kümülatif iptal edilen hacim, gerçekleşen hacim, kalan hacim, kapanış zamanı ve yaşam süresi hesaplandı. Emir ancak kalan miktar sıfıra ulaştığında kapanmış kabul edildi; kısmi iptal ve kısmi gerçekleştirmeler terminal olay sayılmadı.

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



Şekil 2 — Kapanmış emirlerde yaşam süresi ve emir defterindeki konum dağılımları.

Bu adım, spoofing/layering tespiti için temelini oluşturdu. Çünkü “çok kısa yaşayan”, “neredeyse tamamen iptal edilen” veya “gerçekleşmeden kaybolan” bir emri tanımlamak için önce emrin tüm yaşam döngüsünün doğru izlenmesi gerekiyor.

4. Davranışsal adayların oluşturulması

İlk filtreleme katmanında finansal mikro yapı bilgisini doğrudan kullandım. Spoofing adayları için kısa yaşam süresi, yüksek iptal oranı, düşük gerçekleşme oranı ve mutlak veya görünür piyasa derinliğine göre yüksek emir büyüklüğünü birlikte değerlendirdim.

Layering için ise aynı tarafta, çok kısa bir zaman penceresinde ve yakın fiyat bölgesinde birden fazla fiyat seviyesine yayılan yüksek hacimli emir dizilerini aradım. Eşiklerin mümkün olduğunca verinin kendi dağılımından türetilmesi, sabit ve keyfi sınırların etkisini azaltmak açısından önemliydi.

Davranışsal çıktı	Emir sayısı
Spoofing candidate order	7.726
Layering candidate order	2.702
Benzersiz candidate order	9.108

Bu 9.108 emir, analiz edilen kapanmış emirlerin yaklaşık %5'ini oluşturdu. Buradaki “candidate” ifadesi özellikle önemli: bu aşama bir manipülasyon kararı üretmiyor; yalnızca daha ileri analiz için davranışsal olarak ilgi çekici bir havuz oluşturuyor.

5. Emirlerden olaylara: DBSCAN ile episode oluşturma

Spoofing tek bir büyük emirle gerçekleşebilirken layering çoğu zaman birbirine yakın çok sayıda emrin oluşturduğu bir yapıdır. Bu nedenle candidate emirleri doğrudan model gözlemi olarak bırakmak yerine, zaman ve fiyat yakınlığına göre tek bir analitik olay altında topladım.

DBSCAN burada anomali tespit modeli olarak değil, episode oluşturma mekanizması olarak kullanıldı. BUY ve SELL tarafları ayrı kümелendi. Zaman ve fiyat eksenleri piyasa açısından yorumlanabilir ölçeklere

getirildi; $\text{eps}=0,3$ ve $\text{min_samples}=2$ ile ilişkili emirler episode'lara dönüştürüldü.

DBSCAN'in "noise" etiketi finansal anomali anlamına gelmez. Komşusu olmayan bir candidate emir, tek başına da güçlü bir spoofing örneği olabileceği için tek-emirlik episode olarak korunabilir.



Bu işlem sonunda 9.108 candidate emir, 3.231 candidate episode'a dönüştü. Böylece sonraki modelleme aşaması emir satırları yerine ekonomik anlamı daha güçlü olay birimleri üzerinden yürütüldü.

6. Normal piyasa davranışını referans almak

Etiketsiz anomali tespitinde referans dağılımı kritik. Isolation Forest yalnızca behavioral kriterlerden geçmiş candidate episode'larda çalıştırılırsa, model genel piyasadaki sıra dışılığı değil; zaten şüpheli seçilmiş olayların kendi içindeki farklarını ölçer.

Bu nedenle candidate olmayan emirlerden de aynı DBSCAN tanımıyla control episode'lar oluşturdum. Candidate ve control tarafında episode üretim mekanizmasını değiştirmedim. Bu seçim, modelin gerçek piyasa taban oranını daha iyi yansıtan bir referans havuzunda çalışmasını sağladı.

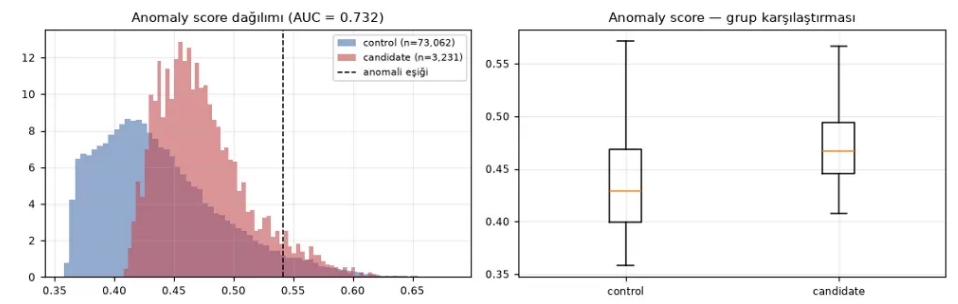
Referans havuzu	Episode sayısı
Candidate	3.231
Control	73.062
Toplam	76.293

7. Isolation Forest ile anomalilik ölçümü

Her episode; emir sayısı, fiyat seviyesi sayısı, toplam hacim, episode süresi, yaşam süresi, iptal ve gerçekleşme oranları, touch distance, spread, order-book imbalance, local volatility ve piyasa aktivitesi gibi özelliklerle temsil edildi.

Candidate/control bilgisi, spoofing/layering tipi ve olay sonrasındaki fiyat değişimleri modele feature olarak verilmedi. Böylece behavioral filtrelerin doğrudan modele sızması ve future leakage engellendi.

Isolation Forest candidate + control ortak havuzunda fit edildi. Candidate episode'ların ortalama anomaly score'u 0,4742, control episode'ların ortalaması 0,4388 olarak bulundu. İki grubun anomaly-score sıralamalarının ayrışmasını özetleyen AUC değeri 0,7322 oldu.



Şekil 3 — Candidate ve control episode'ların anomaly-score dağılımları.

Buradaki AUC, bir classification accuracy değeri değil. Gerçek spoofing/normal etiketleri olmadığı için precision, recall veya F1 gibi supervised metrikler hesaplanamaz. AUC yalnızca iki grubun anomaly-score dağılımlarının ne ölçüde ayrıştığını gösteren yardımcı bir ölçü.

Çalışmanın en güçlü bulgusu, davranışsal kriterlerle oluşturulan candidate havuzunun, bu etiketi hiç görmeyen Isolation Forest tarafından da genel piyasaya göre daha sıradışı skorlanmasıydı.

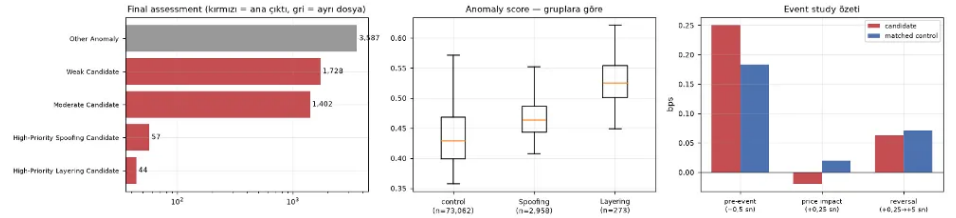
8. Şüpheli olayların önceliklendirilmesi

Final değerlendirmede üç farklı sinyal birlikte ele alındı: Behavioral Evidence, Anomaly Evidence ve Price-Pattern Signal. Behavioral Evidence olayın spoofing/layering karakteristiğini, Anomaly Evidence genel piyasa davranışından istatistiksel sapmayı, Price-Pattern Signal ise olay çevresindeki fiyat örüntüsünün hipotezle uyumunu temsil ediyor.

Price-Pattern Signal özellikle nedensellik kanıtı olarak kullanılmadı. Bir event çevresinde fiyat hareketi görmek, event'in o hareketin nedeni olduğunu göstermez.

Nihai değerlendirme	Episode
High-Priority Spoofing Candidate	57
High-Priority Layering Candidate	44
Moderate Candidate	1.402
Weak Candidate	1.728

Sonuçta 57 spoofing ve 44 layering olmak üzere toplam 101 High-Priority Candidate elde edildi. Bunlar “101 manipülasyon vakası” değil; insan incelemesine öncelikli olarak sunulabilecek olaylar.

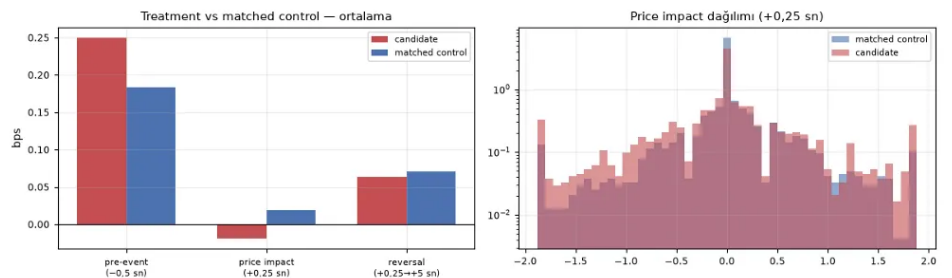


Şekil 4 — Final candidate dağılımı, anomaly-score grupları ve fiyat analizi özetinin birlikte görünümü.

9. Fiyat etkisini ayrı bir hipotez olarak sınamak

Davranışsal ve anomalilik açısından güçlü adaylar bulunduğu doğal olarak bir sonraki soru ortaya çıkıyor: Bu olayların ardından fiyat davranışı normal piyasa olaylarından farklı mı?

Bu soruyu modelin içine gömmek yerine ayrı bir event-study katmanında inceledim. Candidate episode’lar saat, spread, imbalance, volatilité, piyasa aktivitesi, emir sayısı ve toplam hacim gibi koşulları benzer control episode’larla eşleştirildi. 3.231 candidate’ın 3.153’ü için uygun matched control bulundu; eşleşme oranı yaklaşık %97,6 oldu.



Şekil 5 — Candidate ve matched-control gruplarında pre-event, price-impact ve reversal karşılaştırması.

Price impact karşılaştırmasında Wilcoxon testi $p=0,00413$ verdi. Ancak matched rank-biserial effect size $-0,0682$ ile oldukça küçüktü ve farkın yönü spoofing hipotezinin beklediği yönün tersindeydi. Reversal için $p=0,673$ olduğundan anlamlı bir fark bulunmadı.

Daha kritik sonuç pre-event placebo testinden geldi. Olay başlamadan 0,5 saniye önce candidate ve control grupları arasında $p=0,000518$ düzeyinde bir fark vardı. Bu, event sonrasındaki farkın tamamını candidate olaya nedensel olarak atfetmeyi engelliyor.

İstatistiksel olarak anlamlı bir fark, tek başına ekonomik olarak büyük bir etki veya nedensellik anlamına gelmez.

10. Sonuç: Anomali tespiti ile manipülasyon kanıtı aynı şey değil

Bu çalışma, etiketsiz emir defteri verisinde spoofing/layering karakteristiği taşıyan olayların sistematik biçimde daraltılabileceğini gösteriyor. Davranışsal kriterler 9.108 candidate emir üretti; DBSCAN bunları 3.231 episode'a dönüştürdü; Isolation Forest ise candidate episode'ların genel piyasa davranışına göre daha anomali bir dağılım sergilediğini gösterdi.

Finalde 101 yüksek öncelikli inceleme adayı elde edildi.

Ancak matched-control ve placebo analizleri bu olayların fiyatı nedensel olarak kendi lehlerine hareket ettirdiğini desteklemedi.

Bu ayrım bence çalışmanın en önemli tarafı. Çünkü piyasa gözetiminde güçlü bir sistemin görevi her anomalinin arkasına “manipülasyon” etiketi yapıştırmak değil; analistin dikkatini doğru olaylara yönlendirmek ve modelin kanıtlayamayacağı şeyi kanıtlamış gibi göstermemek.

Machine Learning ✓

Data Science ✓

Anomaly Detection +

Algorithmic Trading +

Financial Markets +

**Written by Samet Yorulmaz**

0 followers · 1 following

Edit profile

No responses yet

Samet Yorulmaz

What are your thoughts?

More from Samet Yorulmaz

 Samet Yorulmaz · Jul 17

Tek Şehre Sığın Gökyüzü: Türkiye Havacılığının 18 Yıllık Dönüşümü

DHMI verileriyle 2007–2025 arası bir
mercek.



See all from Samet Yorulmaz

[Help](#) [Status](#) [About](#) [Careers](#) [Press](#) [Blog](#) [Store](#) [Privacy](#) [Rules](#)
[Terms](#) [Text to speech](#)