
**CORAZA WAF LOG ANALİZİ
VE
SOM TABANLI ANOMALİ TESPİT SİSTEMİ**

- Self-Organizing Map Algoritması ile Güvenlik Verilerinin Kümelenmesi ve Tehdit Tespiti

Zübeyir Tosun

İÇERİK

1. Araştırmanın Amacı
2. Araştırmanın Soruları
3. Giriş
4. Yöntem
5. Bulgular ve Sonuç
6. Kaynakça

Günümüz Web Güvenliği Zorlukları

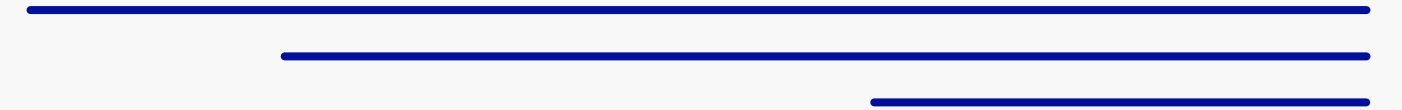


➤ **Veri Hacmi Problemi:** WAF sistemleri günde milyonlarca güvenlik logu üretiyor, manuel analizi imkansız kılıyor.

➤ **Manuel Analiz Zorluğu:** Siber tehditlerin karmaşıklığı ve sayısı arttıkça, insan gücüyle logları incelemek yetersiz kalıyor.

➤ **Yüksek Yanlış Pozitif Oranı:** Geleneksel WAF kuralları, meşru trafiği yanlışlıkla engelliyor, operasyonel yükü artırıyor.
Geleneksel sistemlerde %15-20 yanlış alarm oranları

➤ **Gerçek Zamanlı İhtiyacı:** Saldırıların erken tespiti için hızlı analiz gereksinimi.
Anında tespit ve müdahale, kritik öneme sahip.

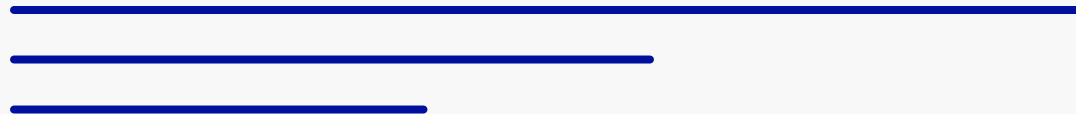


SOM Tabanlı Akıllı Güvenlik Analiz Sistemi



Ana Çözüm Bileşenleri:

- 1 Self-Organizing Map (SOM):** Denetimsiz öğrenme ile log verilerini kümeleyerek benzer saldırı paternlerini gruplandırır.
- 2 Coraza WAF Entegrasyonu:** Modüler ve performanslı Coraza WAF loglarını doğrudan işleyerek gerçek zamanlı analiz sağlar.
- 3 CI/CD Otomasyonu:** Geliştirme ve dağıtım süreçlerini otomatize ederek, sistemin sürekli güncellenmesini sağlar.
- 4 İnteraktif Gösterge Paneli:** Streamlit ile analiz arayüzü sunumu.



SOM Algoritması ve WAF Log Analizi



Self-Organizing Map (SOM), yüksek boyutlu veri kümelerini iki boyutlu bir haritaya indirgeyerek görselleştiren güçlü bir yapay sinir ağıdır.

Yüksek Boyutlu Veri Görselleştirme

WAF loglarındaki karmaşık özellikleri 2D haritada anlaşılır hale getirir.

BMU (Best Matching Unit) ile Kümeleme

Her bir log kaydını, haritadaki en benzer nörona (BMU) atayarak kümelenmeyi sağlar.

Quantization Error ile Anomali Tespiti

BMU'sundan uzak olan veya atanamayan logları potansiyel anormallik olarak işaretler.

Benzer Saldırı Kalıplarını Gruplandırma





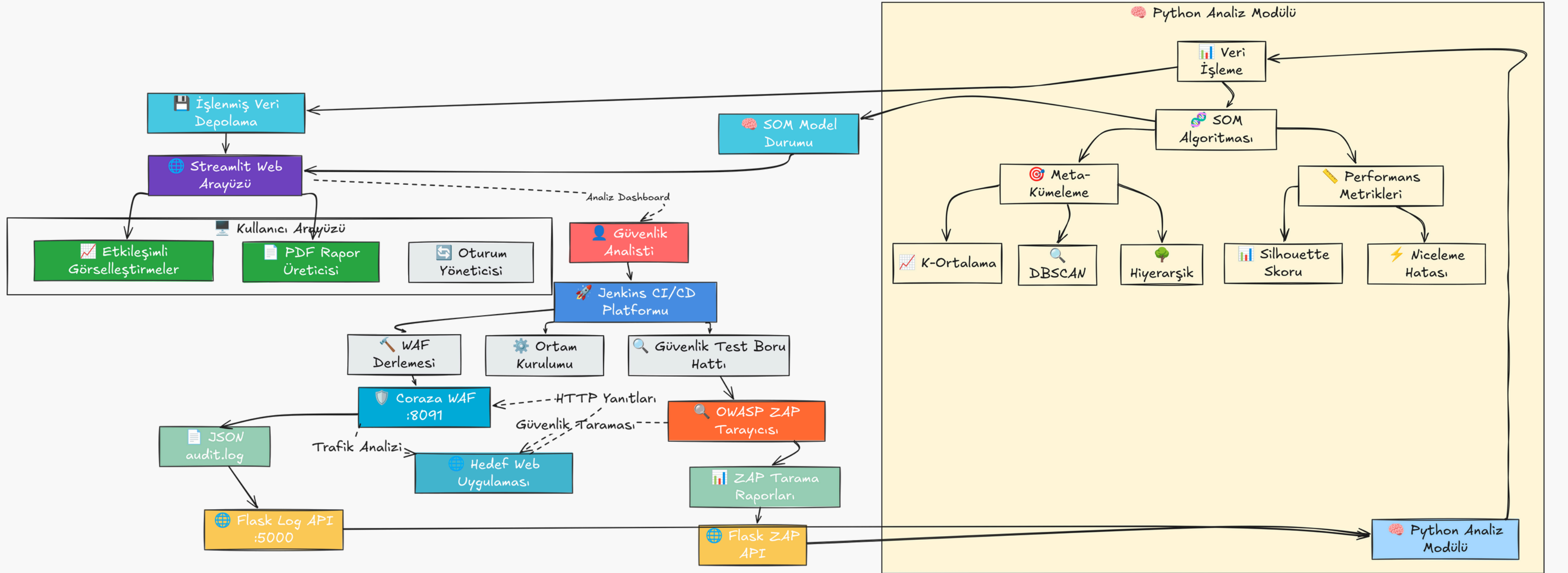
Mikroservis Tabanlı Sistem Mimarisi

Projemiz, modern ve esnek bir yapı sağlamak amacıyla mikroservis mimarisi prensiplerine göre tasarlanmıştır. Bu yapı, her bir bileşenin bağımsız olarak geliştirilmesine, dağıtılmasına ve ölçeklenmesine olanak tanır.

- **Jenkins CI/CD Pipeline:** Otomatik derleme, test ve dağıtım süreçlerini yönetir.
- **Coraza WAF (Go):** Web trafiğini analiz eder ve güvenlik loglarını üretir.
- **OWASP ZAP Güvenlik Taraması:** Uygulama zafiyetlerini otomatik olarak tespit eder.
- **Python Analiz Modülü:** Coraza loglarını işler ve SOM algoritmasını uygulayarak kümeleme yapar.
- **Streamlit İnteraktif Panel:** Analiz sonuçlarını ve anormallikleri gerçek zamanlı olarak görselleştirir.



Mikroservis Tabanlı Sistem Mimarisi



Kullanılan Teknolojiler ve Araçlar

Projemizin geliştirme sürecinde, her bir katman için güncel ve etkili teknolojiler tercih edilmiştir. Bu seçimler, sistemin performansını, ölçeklenebilirliğini ve sürdürülebilirliğini sağlamak üzere yapılmıştır.



Backend

- Go → log üretimi ve HTTP proxy
- Python → analiz motoru ve veri işleme
- Flask → Jenkins ile Python arası köprü



ML/Analytics

- MiniSom → Self-Organizing Map eğitimi
- Scikit-learn → Meta-kümeleme algoritmaları, Performans metrikleri ve Boyut indirgeme
- NumPy → JSON log verilerinin DataFrame'e dönüştürülmesi



Frontend

- Streamlit → Ana web arayüzü ve kullanıcı etkileşimi
- Plotly → İnteraktif SOM haritaları ve serpilme diyagramları
- Matplotlib → İstatistiksel çizimler ve ısı haritaları

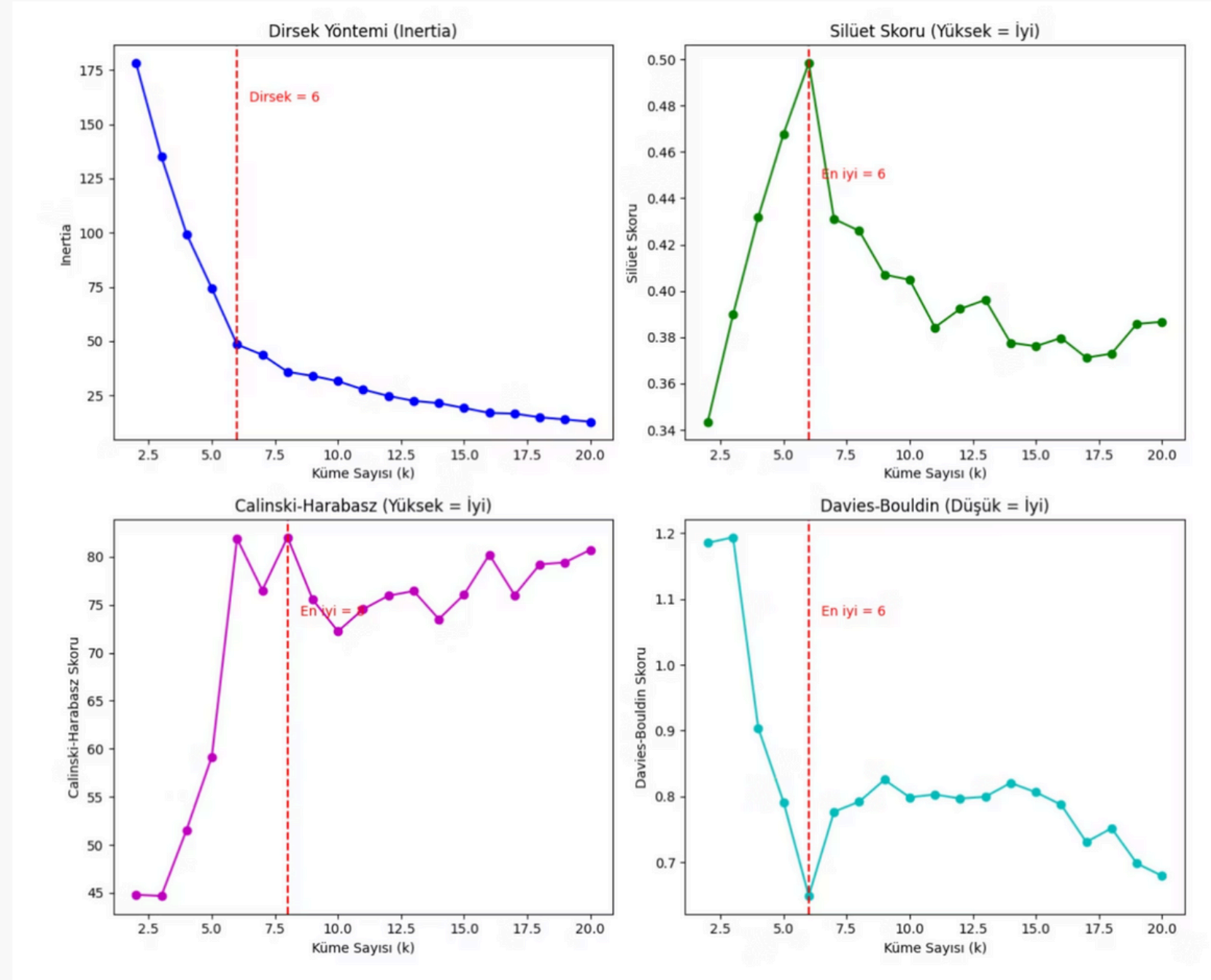


DevOps

- Jenkins → 4 farklı CI/CD pipeline yönetimi
- Docker → İzole test ortamı sağlama, çarpaz platform
- OWASP ZAP → Saldırı simülasyonu ve zafiyeti tetikleme

Kapsamlı Analiz ve Görselleştirme

Sistem, basit anomali tespitinin ötesine geçerek, güvenlik verilerinin derinlemesine analizini ve kullanıcı dostu görselleştirmesini sağlayan gelişmiş özellikler sunmaktadır.



Meta-Kümeleme

SOM tarafından oluşturulan kümelere ek olarak K-means, DBSCAN ve Hiyerarşik Kümeleme gibi yöntemlerle daha fazla detaylandırılır.

Boyut İndirgeme

PCA, t-SNE, UMAP gibi tekniklerle yüksek boyutlu veriler, daha az boyutta görselleştirilir ve yorumlanır.

Optimal Küme Sayısı Belirleme

Dirsek metodu ve silüet analizi gibi teknikler kullanılarak en uygun küme sayısı otomatik olarak belirlenir.

PDF Rapor Üretimi

Analiz sonuçları, periyodik veya isteğe bağlı olarak detaylı PDF raporları şeklinde dışa aktarılabilir.

Gerçek Zamanlı Gösterge Paneli

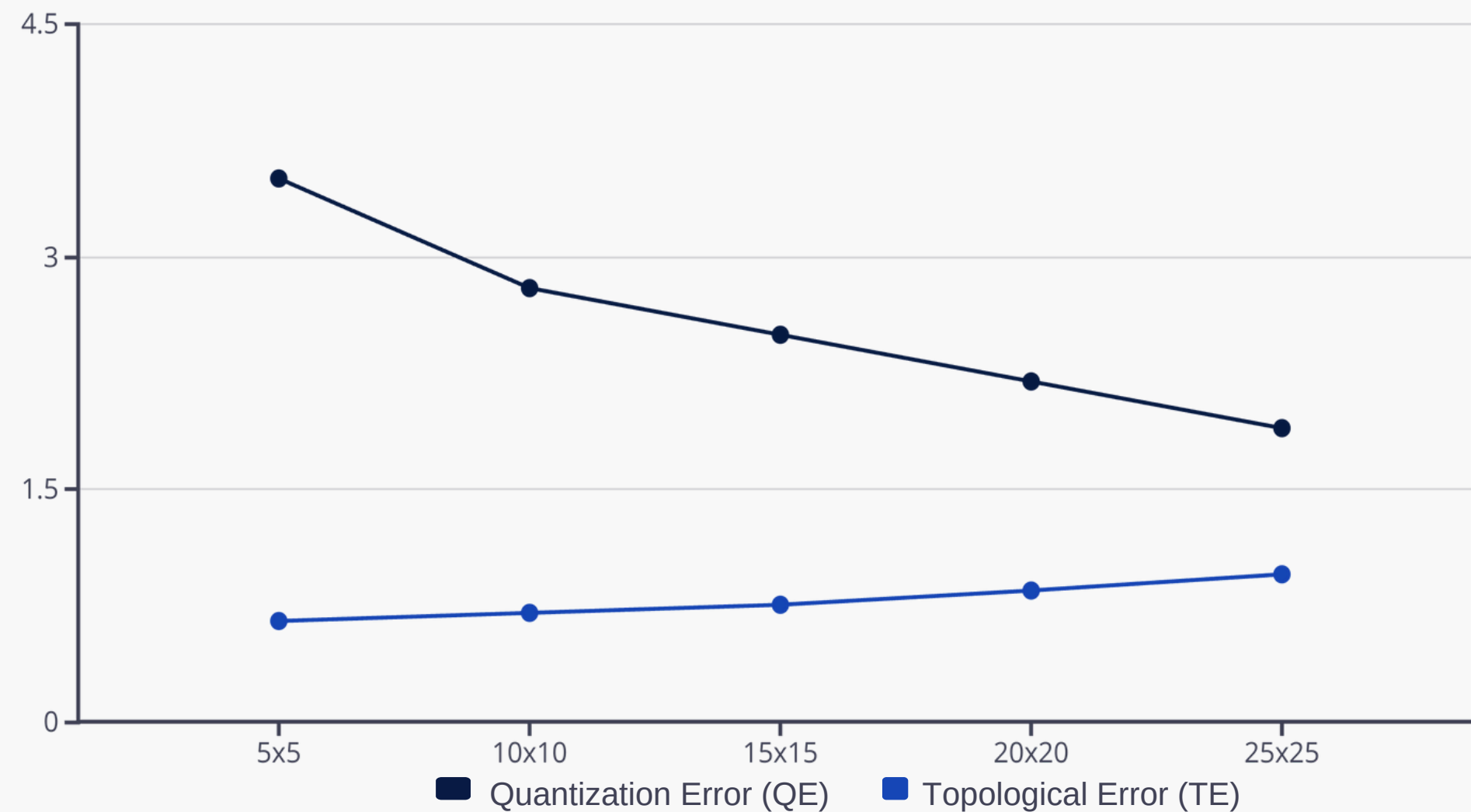
Streamlit tabanlı arayüz ile güvenlik olayları ve anormallikler izlenir.

SOM Algoritması Eğitim Sonuçları

SOM Eğitim Parametreleri

- Grid Boyutları: Geniş bir aralıkta, **5x5'ten 25x25'e** kadar kapsamlı testler yapıldı.
- Öğrenme Oranı: Başlangıçta **0.5** olarak ayarlanıp, iterasyonlar boyunca **0.1'e** kadar kademeli olarak azaldı.
- Komşuluk Fonksiyonu: Nöral ağların komşuluk ilişkilerini tanımlayan **Gaussian** fonksiyonu (sigma=1.0) kullanıldı.
- İterasyon Sayısı: Algoritma yakınsama ve optimal harita oluşumu için **1000 döngü** boyunca eğitildi.

Grid boyutlarının Quantization Error (QE) ve Topological Error (TE) üzerindeki etkisini gösteren aşağıdaki grafik, optimal parametrelerin belirlenmesinde kritik rol oynamıştır. QE, grid boyutu arttıkça düzenli olarak azalırken, TE'de 15x15'ten sonra hafif bir artış gözlenmiştir.



Performans Sonuçları

- En İyi Grid: **10x10 ile 15x15** arasındaki grid boyutları, anomali tespiti için en dengeli performansı sundu.
- Quantization Error: Verilerin SOM haritasına ne kadar iyi yerleştirildiğini gösteren ortalama **2.63** olarak ölçüldü.
- Eğitim Süresi: **10x10'luk bir grid** için eğitim süresi **0.1 saniyenin altında** tamamlandı, hızlı işleme kabiliyeti sağlandı.
- Bellek Kullanımı: Aynı **10x10'luk grid** için bellek tüketimi oldukça düşüktü, sadece **200 MB** olarak kaydedildi.

Sistem Performans Testleri ve Deneysel Bulgular

Test Ortamı

- Donanım: AMD Ryzen 7 7435HS, 16GB 4800Mhz DDR5, 512GB NVMe
- İşletim Sistemi: Pop!_OS 22.04 LTS (Linux Kernel 6.12.10)
- Test Verisi: 2,000 kayıtlık gerçek ZAP Scanner çıktısı

Performans Sonuçları Tablosu

İşleme Süresi	0.08 sn
Bellek Kullanımı	45 MB
CPU Kullanımı	%22

Ana Bulgular:

K-means: En dengeli performans - 2/3 metrikte en iyi
DBSCAN: En iyi küme ayrışması (Silhouette) ama genel performans düşük

Sonuç: K-means WAF log analizi için en uygun algoritma







Meta-Kümeleme Karşılaştırma Tablosu

Algoritma	Silhouette ↑	Calinski-H ↑	Davies-B ↓	Genel Değerlendirme
K-means	0.384	5,035.6 ✓	0.812 ✓	En dengeli ★
DBSCAN	0.589 ✓	98.3	1.0	Küme Kalitesi Yüksek
Hierarchical	0.325	4,280.3	0.9	Orta performans

Proje Başarıları ve Bilimsel Katkıları



Teknik Başarılar

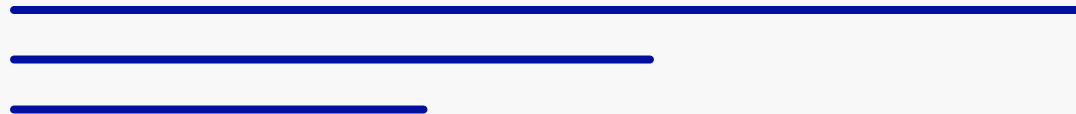
-  Otomatik WAF log kümeleme sistemi
-  CI/CD pipeline entegrasyonu
-  Çoklu JSON format desteği
-  İnteraktif analiz platformu
-  Meta-kümeleme algoritması entegrasyonu
-  PDF rapor üretim sistemi

Bilimsel Katkıları

- Metodolojik: SOM'un güvenlik alanında yeni bir uygulama
- Algoritmik: Adaptif ızgara boyutlandırma formülü
- Teknik: Hibrit anomali tespit yaklaşımı
- Sistem: Mikroservis tabanlı güvenlik analiz mimarisi

Fonksiyonel Doğrulamalar

- Veri yükleme ve işleme: %99 başarı
- SOM eğitimi ve BMU hesaplama: Doğrulandı
- Görselleştirme üretimi: İnteraktif çalışıyor
- Meta-kümeleme analizi: 3 algoritma destekli



Sonuçlar, Katkılar ve Gelecek Perspektifler



"SOM algoritmasının WAF log analizi alanında uygulanabilirliğini göstererek, fonksiyonel ve kullanıcı dostu bir güvenlik analiz prototipi geliştirilmiştir."

Akademik Değer

- Makine öğrenmesi tabanlı güvenlik çözümlerine metodolojik katkı
- SOM algoritmasının yeni uygulama alanı keşfi

Gelecek Vizyon

"Siber güvenlik alanında yapay zeka destekli, proaktif ve otomatik tehdit tespit sistemlerinin yaygınlaştırılması"

Pratik Faydalar

- Güvenlik analistleri için zaman tasarrufu (%80 analiz süresi azalması)
- Otomatik anomali tespiti ile insan hatası olabildiğince azaltma
- İnteraktif görselleştirme ile kolay yorumlanabilir sonuçlar

"Bu çalışma, gelecek araştırmalar için sağlam bir temel oluşturmakta ve gerçek güvenlik analizi performansının değerlendirilmesi için kapsamlı deneysel çalışmalara ihtiyaç olduğunu göstermektedir."



**Dinlediğiniz için ve Zafer SERİN Hocam’a katkılarından dolayı
teşekkür ederim.**

Zübeyir Tosun



Kaynakça:

1. https://youtu.be/H9H6s-x-0YE?si=Scb_sNN7v9TzXZ00 - SOM Anlatım Videosu
2. Kendi Raporum