

ALI HAYDAR KURBAN

151055058

CSE 454 DATA MINING

HOMEWORK1

ANOMALY DETECTION

Article : Survey on Anomaly Detection using Data Mining Techniques



Anomaly Detection Nedir?

- Anomaly Detection, temel anlamıyla bir verideki beklenmedik durumların bulunmasını sağlayan bir tekniktir. Bu beklenmedik durumlar, verinin alışlagelmiş halinden farklı geldiği durumlardır. Beklenmedik durumlara literatürde outliers, exceptions veya anomaliler denilmektedir.

Anomaly Detection Örneđi

- Beklenmedik duruma örnek vermek gerekirse her ay aylık harcaması 100-150 lira olan bir kiři için son ay 2000 liralık bir harcama bedelinin gelmesi bir outlier' dır. Böyle bir durumda bankanız anomaly detection tekniklerini kullanıyorsa , bunun bir kredi kartı dolandırıcılığı (Credit Card Fraud Detection) olabileceđini öngörüp size bir mesaj yollar.



Anomaly Detection Tipleri?

- **Point Anomalies**

- Bireysel bir veri örneği eğer diğer normal verilerden uzaktaysa bu bir anomali veridir.

- **Contextual Anomalies**

- Belirli bir verimiz bazı durumlarda anomaliye işaret ederken diğer durumlarda normal bir veriye işaret ediyorsa bu bir Contextual anomaliye örnektir.

- **Collective Anomalies**

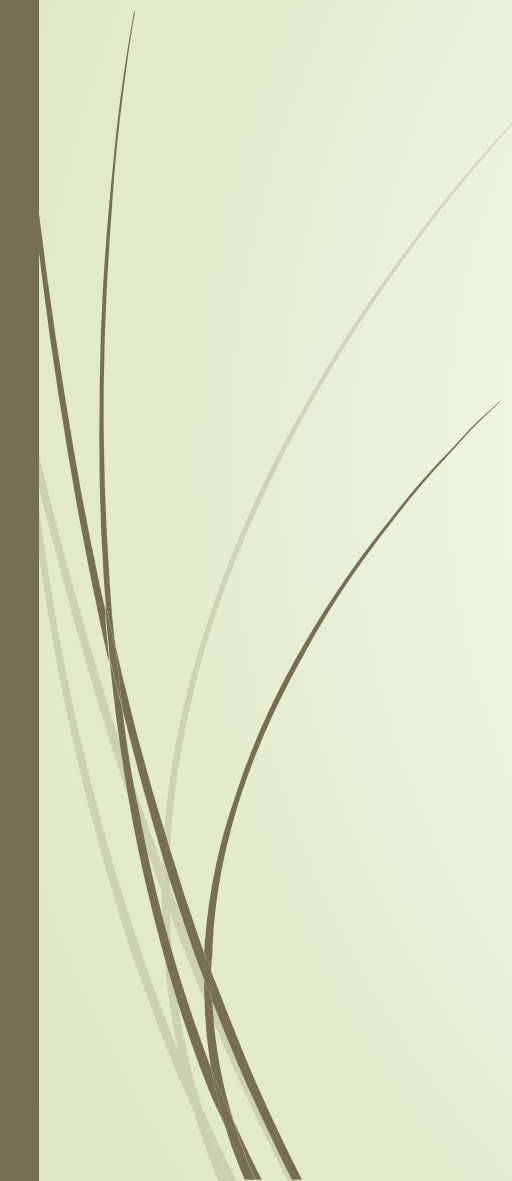
- Birbiriyle ilişkili olan veriler tüm verisetinde anomali davranış oluşturuyorsa bu bir collective anomaliye örnektir.

Anomali Detection' da Karşılaşılan Zorluklar

- Bir normal veri anomali verilerin bulunduğu kümelenmeye, bir anomali veri ise normal verilerin bulunduğu kümelenmeye yakın olabilir. Bu durumda da anomali tespiti bir hayli zorlaşır.
- Normal diye nitelendirdiğimiz davranışlar veya veriler zaman içinde değişime uğrayabilir.
- Belli bir anomali tespiti tekniğini her alana uygulamak mümkün olmayabilir.
- Verisetlerindeki gürültü “noise” anomalilerin tespit edilmesi için ciddi bir çalışmayla temizlenmesi gereklidir. Ancak gürültülerin ayırt edilmesi oldukça zor bir süreçtir.



Data Mining Teknikleri ile Anomaly Detection

- CLASSIFICATION BASED ANOMALY DETECTION TECHNIQUES
 - NEAREST NEIGHBOR BASED ANOMALY DETECTION TECHNIQUES
 - CLUSTERING BASED ANOMALY DETECTION TECHNIQUES
- 

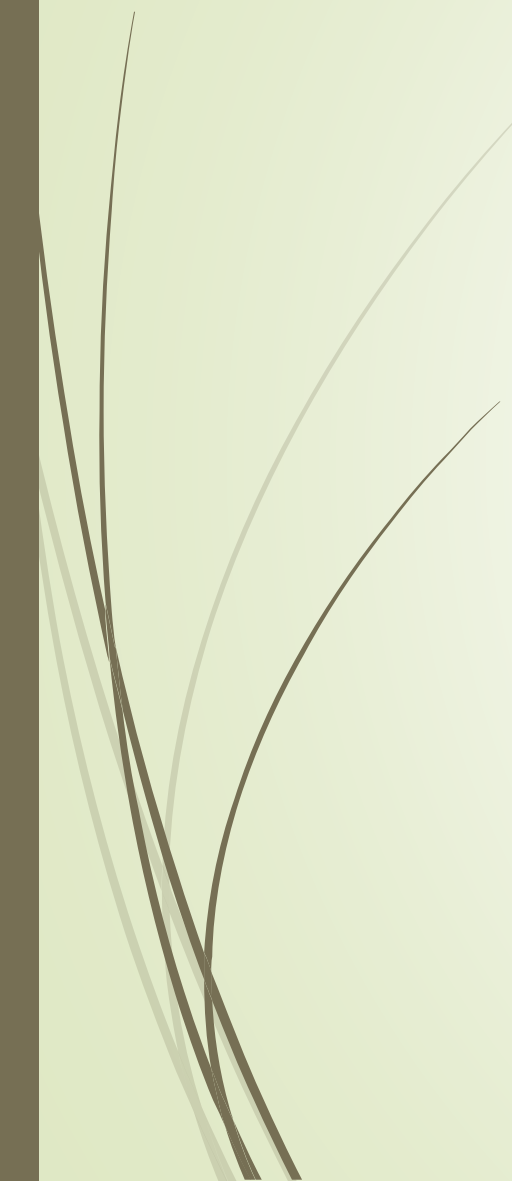


CLASSIFICATION BASED ANOMALY DETECTION TECHNIQUES

- Sınıflandırma, bir dizi etiketli veri örneğinden (eğitim) bir model öğrenmek ve daha sonra öğrenilmiş modeli (test) kullanarak bir test örneğini sınıflardan birine dahil etmek için kullanılır.
- Sınıflandırma tabanlı anomali tespit teknikleri benzer iki fazlı bir şekilde çalışır. Eğitim aşaması, mevcut etiketli eğitim verilerini kullanarak bir sınıflandırıcıyı öğrenir. Test aşaması Bir test örneğini sınıflandırıcıyı kullanarak normal veya anormal olarak sınıflandırır.



CLASSIFICATION BASED ANOMALY DETECTION TECHNIQUES

- Neural Networks Based
 - Bayesian Networks Based
 - Support Vector Machines Based
 - Rule Based
- 



NEAREST NEIGHBOR BASED ANOMALY DETECTION TECHNIQUES

- En yakın komşu tabanlı anomali tespit teknikleri, iki veri örneği arasında tanımlanan bir mesafe veya benzerlik ölçüsü gerektirir. İki veri örneği arasındaki mesafe (veya benzerlik) farklı şekillerde hesaplanabilir. Öklid, manhattan ve minkowski uzaklık ölçümünde kullanılabilir. Çok değişkenli veri örnekleri için, her bir özellik için uzaklık veya benzerlik genellikle hesaplanır ve daha sonra birleştirilir.



NEAREST NEIGHBOR BASED ANOMALY DETECTION TECHNIQUES

- Using Distance to kth Nearest Neighbor
 - Using Relative Density
- 



CLUSTERING BASED ANOMALY DETECTION TECHNIQUES

- Kümeleme, benzer veri örneklerini kümeler halinde gruplandırmak için kullanılır. Kümelenme öncelikle denetimsiz bir tekniktir.
- Kümeleme temelli anomali tespit teknikleri üç grupta toplanabilir.
- 1) Normal veri örnekleri, verilerdeki bir kümeye aittir.
- 2) Herhangi bir kümeye ait değildir.
- 3) Normal veri örnekleri en yakın küme merkezlerine yakındır ama anomaliler en yakın kümelenme merkezlerinden uzaktır. Normal veri örnekleri büyük ve yoğun kümelere ait iken, anomaliler ya küçük ya da seyrek kümelere aittir.



CLUSTERING BASED ANOMALY DETECTION TECHNIQUES

- k-Means
 - k-Medoids
 - Outlier Detection Algorithms
- 