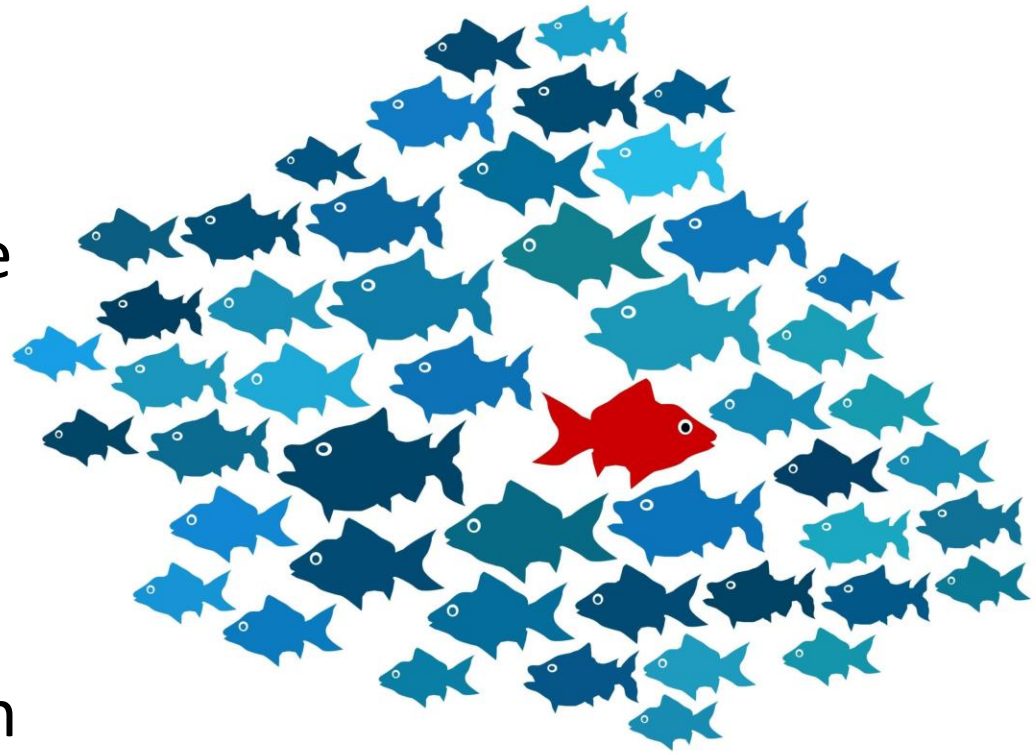
Introduction to Anomaly Detection

Dr. Amit Ranjan

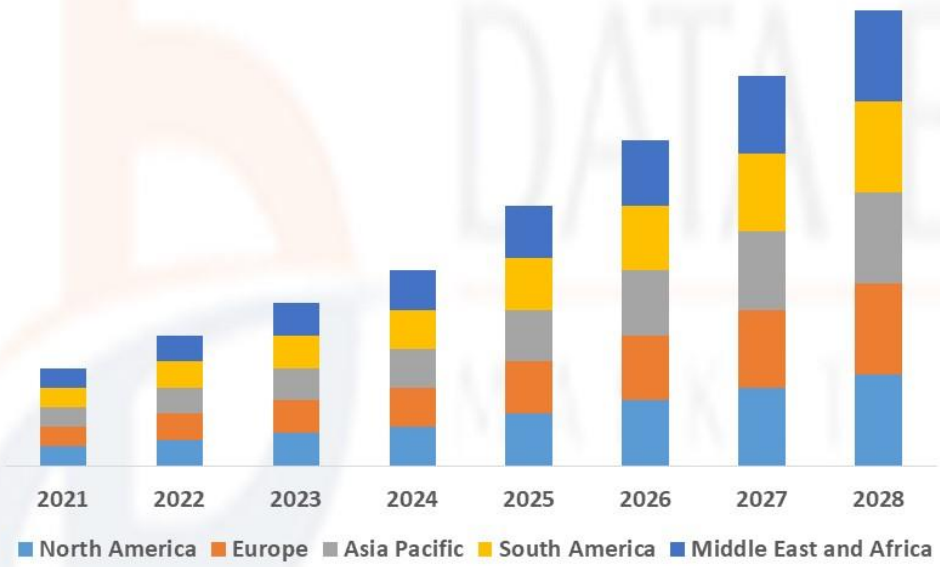
School of Computer Science

What is Anomaly Detection?

- Anomaly Detection is the technique of identifying rare events or observations which can raise suspicions by being statistically different from the rest of the observations.
- These anomalies often signal potential problems like credit card fraud, server failures, or cyber-attacks.
- For example, detecting a sudden spike in online transactions might help prevent fraudulent activities.



Global Anomaly Detection for Professional Market is Expected to Account for USD 16,012.82 Million by 2028

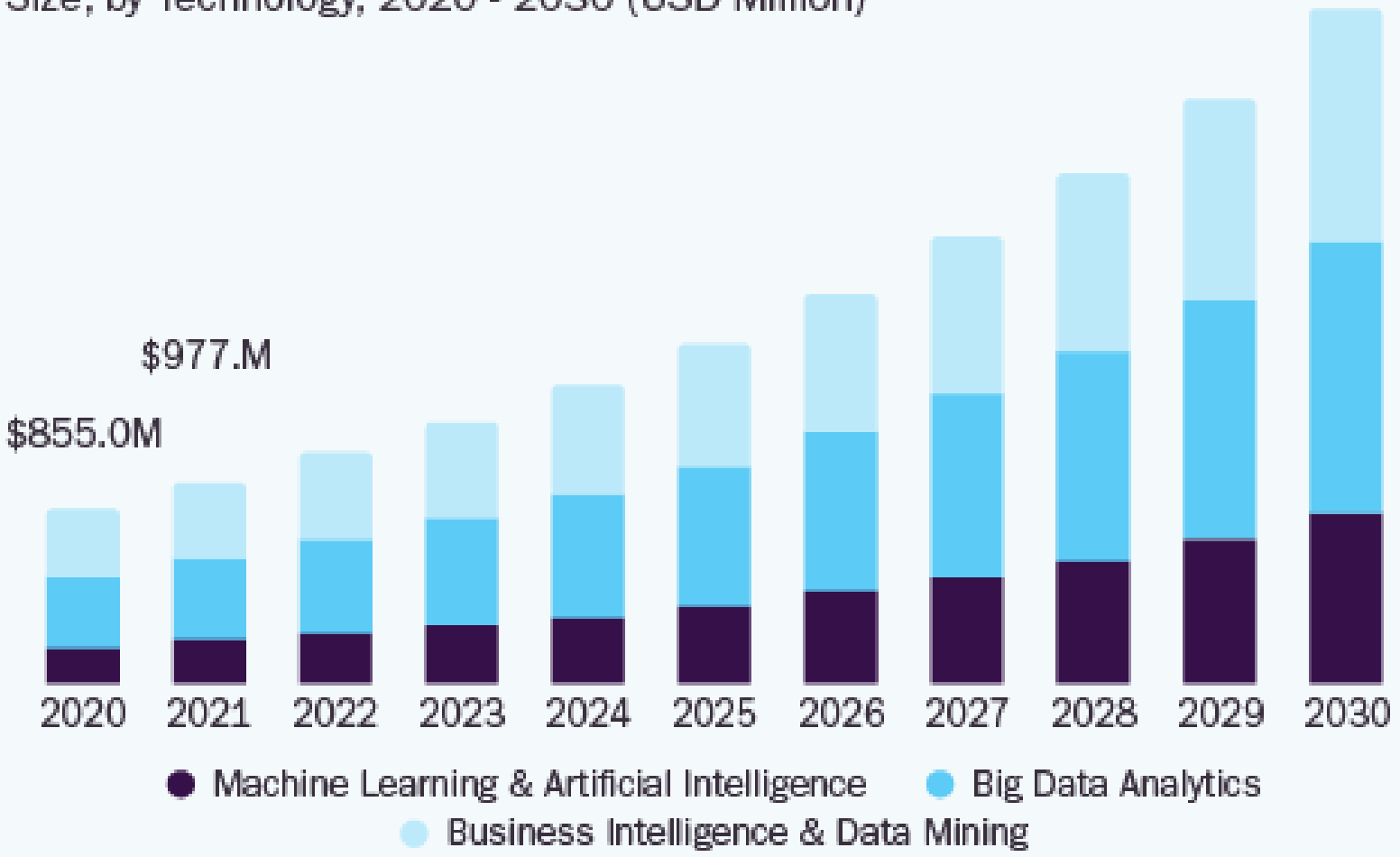


Global Anomaly Detection for Professional Market, By Regions, 2021 to 2028

DATA BRIDGE MARKET RESEARCH

U.S. Anomaly Detection Market

Size, by Technology, 2020 - 2030 (USD Million)



GRAND VIEW RESEARCH

14.3%

U.S. Market CAGR, 2023 - 2030

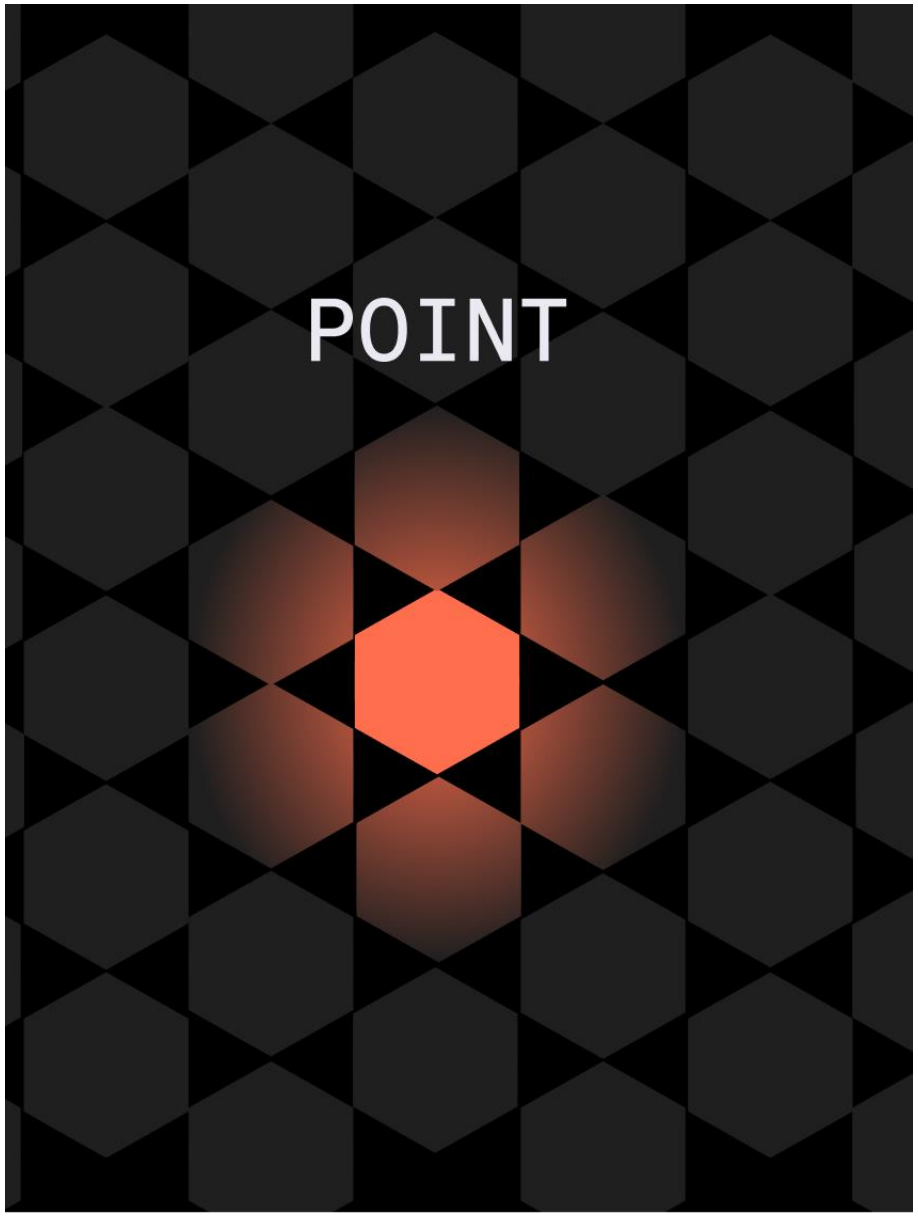
Source:
www.grandviewresearch.com

Types of Anomalies

1. **Point Anomalies**:

Point anomalies, also known as global outliers, are individual data points that exist far outside the rest of the data set.

They can be either intentional or unintentional and may result from errors, noise or unique occurrences. An example of a point anomaly is a bank account withdrawal that is significantly larger than any of the user's previous withdrawals.



Types of Anomalies

2. **Contextual Anomalies**:

Contextual anomalies are data points that deviate from the norm within a specific context. These anomalies are not necessarily outliers when considered in isolation but become anomalous when viewed within their specific context.

For example, consider home energy usage. If there is a sudden increase in energy consumption at midday when no family members are typically home, the anomaly would be contextual. This data point might not be an outlier when compared to energy usage in the morning or evening (when people are usually home), but it is anomalous in the context of the time of day it occurs.

Types of Anomalies

3. C**ollective Anomalies**:

Collective anomalies involve a set of data instances that together deviate from the norm, even though individual instances may appear normal. An example of this type of anomaly would be a network traffic data set that shows a sudden surge in traffic from multiple IP addresses at the same time.

Practical Use Cases of Anomaly Detection

1. Credit card fraud analysis using data mining technology

In today's world, we are literally on the express train to a cashless society. According to the World Payments Report, total non-cash transactions in 2016 totalled 482.6 billion, up 10.1% from 2015. It's huge! Cashless transactions are also expected to grow steadily in the coming years.

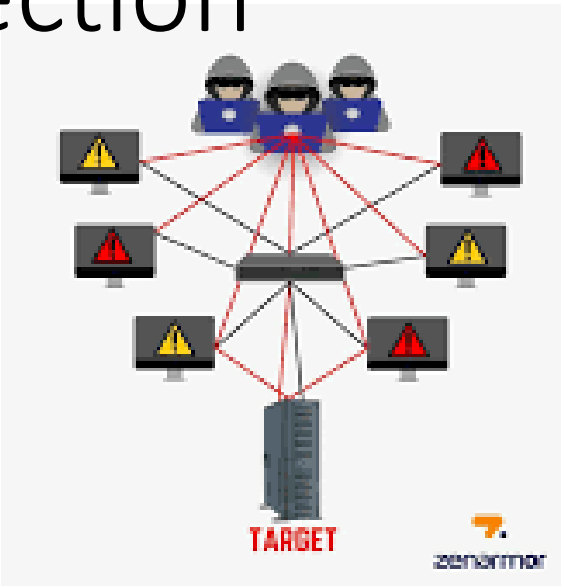
Even with EMV smart chips implemented, it is good to use anomaly detection for a huge number of fraudulent transactions.

So our data scientists are trying to find one of the best solutions for building models that predict fraudulent transactions.

Practical Use Cases of Anomaly Detection

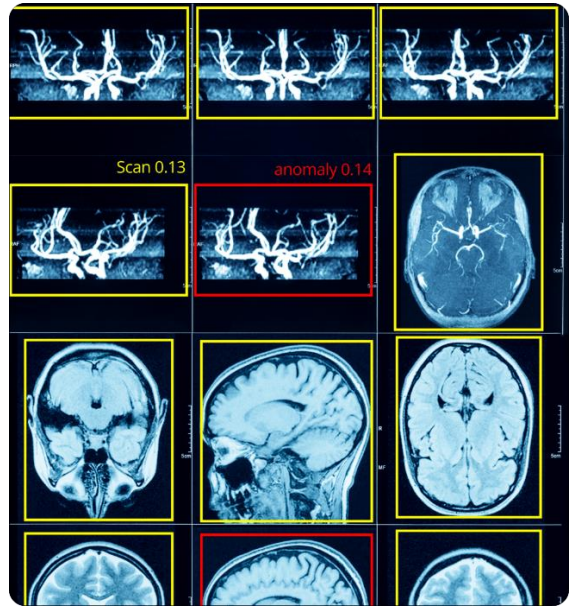
2. Network intrusion detection

Anomaly detection can be used to identify unusual network traffic patterns, i.e. making intrusion detection systems that may indicate an attempted cyber-attack.



3. Medical diagnosis

Anomaly detection can be used to identify unusual patterns in patient data, such as vital signs or test results, that may indicate a medical condition.



Practical Use Cases of Anomaly Detection

4. Manufacturing quality control

Anomaly detection can be used to identify unusual patterns in production data that may indicate a problem with the manufacturing process.

5. Traffic prediction

Anomaly detection can identify unusual traffic data patterns that may indicate an issue with the transportation network, such as a road closure or accident.

6. Environmental monitoring

Anomaly detection can be used to identify unusual patterns in environmental data, such as temperature or air quality measurements, that may indicate a problem.

Calculating Anomalies Over Time

- According to IBM, anomaly detection is calculated over time by comparing current data points to a historical baseline established from past data.
- Using statistical methods or machine learning algorithms to identify significant deviations from the expected pattern.
- By setting thresholds on key metrics and flagging data points that fall outside those boundaries; this process is continuously updated as new data becomes available, allowing the system to adapt to changing trends and better identify anomalies over time.