Seminar Presentation

Title: Isolation-Based Anomaly Detection

Author: FEI TONY LIU and KAI MING TING

Instructor: Prof. Vishnu Narayanan

Priyansu(23N0464)

- ➤ Problem Overview

- ➤ Background

- ➤ Proposed Method

- ➤ Experiments and Results

- ➤ Future Work

- ➤ References

What is Anomaly Detection?

Anomaly Detection is the process of identifying rare events or outliers that deviate significantly from the norm in a dataset. These anomalies can signal potential issues or interesting patterns in the data.

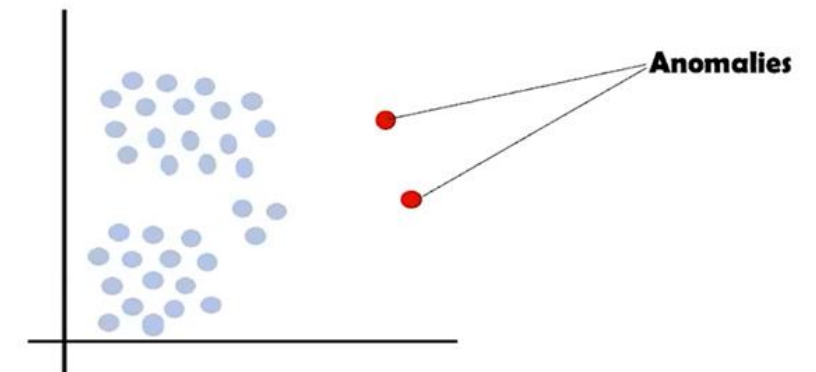## 1). Cybersecurity

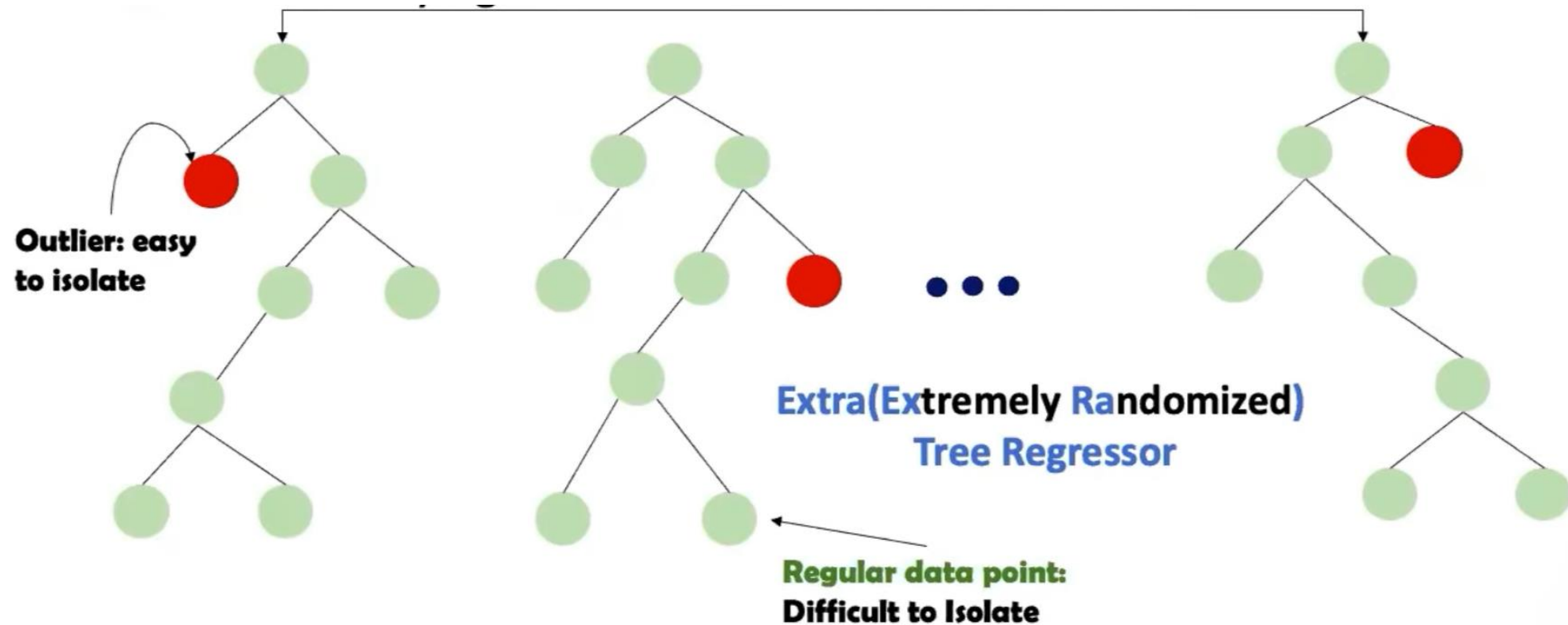## 2).Fraud Detection

3).Health Care

Is there any way that we can looked the data and identify the anomaly , more importantly that we can apply algorithm and detect the outlier

Lets just look the data and trying to understand what is this anomaly between data

Isolation Forest:-Isolation Forest is an anomaly detection algorithm that is particularly effective in identifying outliers or anomalies in a dataset.

Outlier: easy to isolate

Extra(Extremely Randomized) Tree Regressor

Regular data point: Difficult to Isolate

Calculations

➢ We will compute anomaly score for every data point

➢ If it crosses specific threshold we will consider as outlier

Mathematical Formula:

$$S(x, m) = 2^{-E(h(x))/c(m)}$$

If $E(h(x)) \ll c(m)$ ➡ $S(x, m) \approx 1$

If $E(h(x)) \gg c(m)$ ➡ $S(x, m) \approx 0$

If $E(h(x)) = c(m)$ ➡ $S(x, m) = 0.5$

Where, m = Number of points

x = Data Point

E(h(x)) represents the average path length for isolating data point x in a tree.

c(m) represents the average depth of data points in a tree.

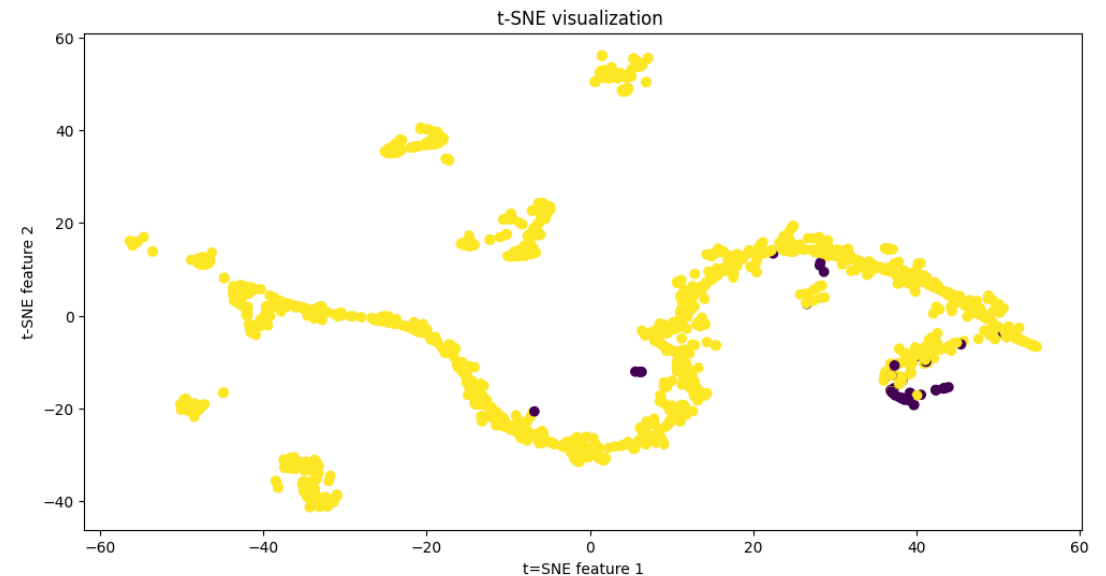Transaction dataset:-This dataset contains rows of known fraud and valid transactions made over Ethereum, a type of cryptocurrency. The shape of the data is (9841,51)
I use the library IsolationForest from sk.learn

Visualization on training data
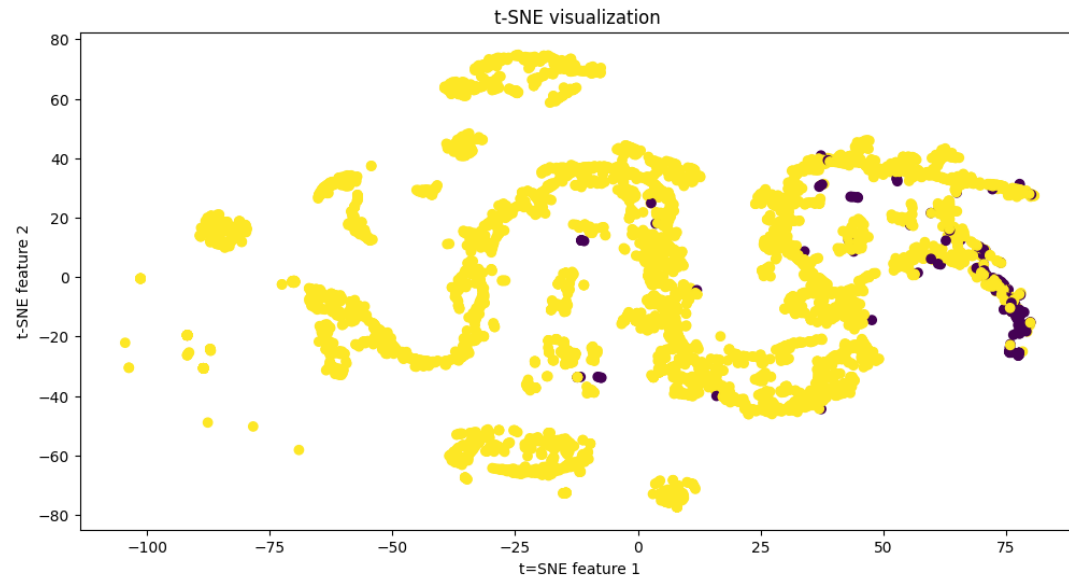
Visualization on testing data

We can use the five of the most popular **outlier detection methods** .

- Tukey's IQR method

- Standard deviation method

- Z-score method

- Modified z-score

- DBSCAN - Density-Based Spatial Clustering of Applications with Noise

[1] https://www.analyticsvidhya.com/blog/2021/07/anomaly-detection-using-isolation-forest-a-complete-guide/

[2] https://github.com/scikit-learn/scikit-learn/blob/872124551/sklearn/ensemble/_iforest.py#L28

[3] https://www.kaggle.com/datasets/vagifa/ethereum-frauddetection-dataset?select=transaction_dataset.csv

[4] https://dl.acm.org/doi/pdf/10.1145/2133360.2133363

Thank you