



Contribution ID: 33

Type: **not specified**

Outlier detection using robust random cut forest

Tuesday, 18 May 2021 10:00 (20 minutes)

Due to the development of sensor devices and ubiquitous computing, we generate an enormous amount of data every second of every day. With access to a gigantic amount of information, it is imperative to analyze it, monitor it, and interpret it correctly so that business decisions are made correctly. When it comes to security, finding the anomalies is only the first step in data analysis. Assessing if the anomaly is really a security threat and understanding the main cause of the anomaly is the answer to a real solution. Therefore, anomaly detection is one of the hottest data science topics that attract researchers in many different fields.

Applications of anomalies may occur in numerous areas, including fraud detection, finance, environmental monitoring, e-commerce, network intrusion detection, medical diagnosis, or social media, among others. Although many anomaly detection algorithms exist for batch setting data, anomaly detection for streaming data nowadays became more popular due to the volume and dynamics of the streams. In this study, we examine the Robust Random Cut Forest (RRCF) technique which was proposed for anomaly detection for streaming data sets. The objective of this study is to study the similarities and differences of this method for batch and stream data settings, assess the performance of the method based on the different type of outliers such as subsequent or point outliers, and compare the performance of this method with some of the “state-of-the-art” algorithms for both settings.

Primary authors: Dr USMAN, Basiru (NC State University); Dr BILLOR, Nedret (Auburn University)

Presenter: Dr USMAN, Basiru (NC State University)

Session Classification: Analytical methods

Track Classification: Data Science in Process Industries