Project: Explaining anomalies in event stream monitoring

1. Project Description
This paper provides high-quality explanations for anomalous behaviours that users annotate on event monitoring results. Given the requirements for explanations, namely, conciseness, consistency with human interpretation, and prediction power, most existing techniques cannot produce explanations that satisfy all three of them. The key technical contributions of this work include a formal definition of optimally explaining anomalies in event monitoring, and three key techniques for generating sufficient feature space, characterizing the contribution of each feature to the explanation, and selecting a small subset of features as the optimal explanation, respectively.

2. Reading Material
Please read the paper from the beginning up to Section 5.4 to understand the relevant algorithms. You may skip “Identifying related partitions”, “Partition alignment” and “Interval labelling” in Section 5.2. Please read Section 6 for the evaluation methodology.

3. Algorithms to Implement
Please implement the following algorithms:
   Section 4: single feature reward (Section 4)
   Section 5: constructing explanations (Section 5.1 and 5.3)

4. Results to Produce
Dataset: event streams will be provided by the course staff.

Results to produce:

- **Fig 15**: Given some anomalies, please reproduce the plot for “XStream” and “XStream-cluster”. There is no need to produce other techniques for comparison.

- **Fig 16**: Please reproduce the plot for “XStream” and “XStream-cluster”. A test set will be provided by the course staff.

- **Fig 17**: Please report the number of features selected in the explanations using the entropy measure. There is no need to produce other techniques for comparison.
5. Ideas for Extension (20%)

This project also requires an extension. We offer a few ideas below. Please feel free to choose one idea to explore in your project. Students are also welcome to propose their own idea.

1) Write an anomaly detection algorithm and detect other anomalies in the dataset.

2) Filter incidentally correlated features. It is known that the explanation construction algorithms may choose features that are incidentally correlated to the small set of labelled anomalous instances. For example, timestamp may be chosen to construct an explanation, but it is actually a bad feature to use because the explanation constructed using timestamp does not generalize to other anomalous instances. Please propose a technique to filter such incidentally correlated features from the constructed explanations.