**Event Log for Incident Management Process**

**Dataset**

Number of instances: 141,712

Number of Features: 36

Link: <https://drive.google.com/file/d/1Qzn5vJgFQpYIqC-5o4f2x2wtGh30uywO/view?usp=sharing>

Description: This is an event log for an incident management process extracted from an audit system in Brazil. The event log is enriched with data loaded from a relational database underlying a corresponding process-aware information system. The additional information of the features is attached at the end of the document.

This project is divided into 3 deliverables, Preprocessing & EDA, Prediction & Clustering and Final Report. For all deliverables you are required to make a report that you will compile and submit in Deliverable 3 i.e., Final Report. The weightage for each deliverable is also given alongside each of them.

**D2: Prediction and Clustering**

1. Clustering:

* Implementation of appropriate clustering technique on features with meaningful results. Hint. One of them may or may not include ‘closed\_code’ and ‘resolved\_by’ fields.
* Perform Outlier Analysis using appropriate fields and explain your findings.

1. Prediction:

* You are required to remove rows where priority is set to be medium and then predict the urgency and priority of the given event.
* Split your data into 80/20 ratio for training and testing data.
* Remove the Class Imbalance Ratio using K-fold Cross Validation.
* Use appropriate fields as your input features (Reminder your data must be numerical values).
* Predict the output for incident features, targeting “urgency” and “priority” using any optimizer.
* Evaluate your results on test data and provide a confusion matrix.
* Also report Precision, Recall, Accuracy and F1 score.