

# **MULTivariate Time Series ANomaly deTection**

Framework (MUTANT)

## **Master Thesis**

at Friedrich-Alexander-Universität Erlangen-Nürnberg  
at the Masters in Data Science in the Data Science Department  
**Image Data Exploration and Analysis (IDEA) Lab**

Principal Supervisor:  
Associate Supervisor:

Prof. Dr. Bernhard Kainz  
Mrs. Johanna Müller

Author:

MD FIROZUR RAHMAN  
Bahnstraße  
91058 Erlangen  
+49 17658574913  
firozur.rahman@fau.de  
22975954

Submission:

6th July 2024

---

### *key features of MUTANT*

- (1) Integrates Graph Convolutional Networks with Variational Autoencoders.
- (2) Validation data, dimension, batch size, shape.
- (3) Scaling and clustering data using k-means.
- (4) Construct graphs by identifying similar and dissimilar nodes based on a feature matrix.
- (5) Create and train the model.
- (6) Use LSTM for store log data.
- (7) test new data using the generated model.
- (8) Calculate the loss function.
- (9) Calculate metrics such as F1 score, precision, recall, and confusion matrix components (TP, TN, FP, FN).
- (10) Functions for point-to-point evaluation.

### *Updated key features of MUTANT (experiments and code extensions)*

- (1) Implemented error handling and exception management to ensure the code runs smoothly.
- (2) Optimizing Functionality and Eliminating Redundancy Using Sophisticated Libraries
- (3) Explore Number of clustering and shape of data.
- (4) Build unique grouping based on clustering use KMeans
- (5) Self-Supervised learning based on clustering
- (6) Produces more user-friendly output and comments on each method's functionality

### Short description of experiments and extensions:

To enhance the MUTANT framework, self-supervised learning is integrated to classify item types in addition to detecting anomalies. This is achieved by using clustering techniques like KMeans with PCA for dimensionality reduction. Future updates could include advanced clustering algorithms, additional self-supervised methods, and optimizations for handling larger datasets. .

#### 0.1 Testing and training result.

```
(base) frz7@frz7 MUTANT % python main.py
Loading data of: SMAP
Data normalized
Train set shape: (135183, 25)
Test set shape: (427617, 25)
Test label shape: (427617,)
test_data: (277952, 25)
val_data: (149665, 25)

***** Final result of Test data *****

Dataset-Name: SMAP
Threshold-value: 99.2500000000032
True Positive: 41805
False Positive: 1855
False Negative: 613
precision: 0.9575125971237946
recall: 0.9855485876313006
f_score: 0.9713233309601832
```

#### 0.2 Testing result of Self-supervised learning

```
Performing brute-force search for the best threshold...
Performing clustering on test data...
Data reshaped for clustering: (5558640, 25)
Data shape: (5558640, 25)
Subsampling data to 10000 points.
Subsampled data shape: (10000, 25)
Applying PCA to reduce to 10 components.
Data shape after PCA: (10000, 10)
Number of clusters determined: 50
Starting KMeans fitting...
KMeans fitting completed.

Number of groups: 50
Group names and counts: {'Group0': 3971, ..... 'Group9': 625}

(base) frz7@frz7 MUTANT %
```