Assignment 2: Anomaly Detection Using LSTM Autoencoders

In this assignment, you will design and implement two different LSTM-based autoencoder models to detect anomalies in time-series or structured data. Anomalies are defined as unusual patterns or outliers that may indicate system faults, unexpected events, or other irregular behaviors. These anomalies are represented in the datasets by values that deviate from normal operating ranges.

In addition to your two LSTM-based models, you are also required to select and implement one state-of-the-art baseline model for comparison. This third model can be another deep learning method or a classical machine learning approach such as K-Means clustering, DBSCAN, or any other well-justified technique suitable for anomaly detection.

Regarding dataset, you are free to choose any dataset (provided below) for this assignment. Working with a single dataset is sufficient to meet the requirements. However, if you are interested in exploring further, you are encouraged to apply the models to additional datasets and include a comparative analysis in your report.

Below are several datasets suitable for anomaly detection tasks:

a. NASA Anomaly Detection Dataset (SMAP & MSL)

This dataset contains telemetry data from the **Soil Moisture Active Passive (SMAP)** satellite and the **Mars Science Laboratory (MSL)** rover. It is commonly used for benchmarking anomaly detection in space systems.

View on Kaggle

b. Network Intrusion Anomaly Detection

This dataset contains network traffic logs and is useful for detecting abnormal activity, such as intrusions or cyber-attacks.

View on Kaggle

c. Financial Transaction Anomalies

Focused on identifying suspicious financial transactions, this dataset simulates real-world fraud detection scenarios using structured data.

View on Kaggle

Report:

Your assignment should include:

1. Data preprocessing (e.g., normalization, handling missing values)
2. Model design, implementation and explanation of two different approaches
3. Training and evaluation of both models using metrics such as reconstruction error, precision, recall, F1-score, or ROC-AUC
4. Visualization of results (e.g., anomaly score plots, confusion matrix, thresholding curve)
5. Comparison and discussion of model behavior, strengths/weaknesses, and real-world applicability
6. A clearly written report with explanations of all decisions, results, and conclusions
7. Instructions on how to reproduce your results

Submission:

Submit a single .zip file containing your report and all implementation files via the Canvas course page. Do not include the dataset in your submission. Name your zip file using your first and last name in the following format:

firstname_lastname.zip

For example: john_doe.zip

Submission Deadline:

Submission deadline is 15th May 2025 until 23:59.

Evaluation

The assignment is graded between A-F. It is important to note that this is an individual assignment. This means that everything that you submit for grading must be created by you. **Plagiarism is not allowed in any form.**