# Assignment1: Anomaly Air Quality Detection using Isolation Forest

## 1. Objective

The primary goal of this assignment is to implement the Isolation Forest algorithm and apply it to identify anomalous samples within the provided "Air Quality" dataset. You are required to build a model that effectively distinguishes outliers of abnormal Air condition from the normal one, and analyze the difference.

## 2. Tasks and Requirements

- **Implementation and Application:**

  Implement the Isolation Forest model. You may use a library like Scikit-Learn or build it from scratch. And then apply the trained model to the "Air Quality" dataset. Your analysis should:

  - Identify the samples classified as anomalies and plot the anomaly_score-sample_index figure.

  - Report and compare the characteristics of the identified anomalous samples against the normal samples.

- **Discussion: Hyperparameter Tuning**

  Explain your methodology for choosing key hyperparameters such as `contamination`, `n_estimators`, and `max_samples`. Justify your selections and analyze how they influenced the model's performance.

## 3. Report and Submission

Your Report should include:

1. the anomaly_score-sample_index figure;

2. the analysis of the characteristics that identified anomalous samples against the normal samples;

3. the discussion of Hyperparameter Tuning;

You should submit:

- all codes;

- your report (a separate PDF file).


**Scikit tools:** IsolationForest — scikit-learn 1.7.2 documentation

**Dataset Detail:**  Air Quality - UCI Machine Learning Repository