

Seminar 6 (Week 6)

Q1. Please read about PyOD library from this link (<https://pyod.readthedocs.io/en/latest/>). The library is used to detect the outliers or anomaly. Please complete the following tasks:

- I. Please generate synthetic data using `pyod.utils.data.generate_data()` utility function. The training and testing data should have 2 features. The training set should have 400 samples and the testing set should have 150 samples.
- II. Please implement four outlier detection algorithms from the pyod library. They are: KNN, Angle based outlier detection, Histogram based outlier detection and Isolation Forest algorithm.
- III. Please write a script to graphically present the outlier for each algorithm. Also present the ROC and precision score for each algorithm.

Q2. Please perform the following activities on the Forex Fundamentals News dataset from 2010 to 2023. This is a time series dataset which provides critical indicators such as ISM Manufacturing and Services PMI, Housing Starts, Non-Farm Employment Change, Unemployment Rate, Consumer Price Index (CPI), Producer Price Index (PPI), and Retail Sales.

- I. Import the dataset into pandas dataframe.
- II. Transform non-numerical data types into numerical. Please focus on the special characters in some of the feature columns and consider to replace them wisely.
- III. Please create individual scatter plots for all the features based on time (Time vs column value)
- IV. Please train a KNN based outlier model to label the anomalies in the dataset.
- V. Please present the anomalies using scatter plot.