

Question 1.1 : What experience on improving the performance of machine learning-based systems does exist in terms of addressing the anomalies (rare event with significant consequences), which may pose considerable impact on technological and/or economic performance of the power system ? How should we distinct anomalies and data outliers between each other ?

Clean and sufficient data in every datasets play a crucial role to improve the performance of Machine Learning so that it can be learned and addressed the anomaly successfully. Therefore, we need to generate our own datasets by simulating the missing incidents ; when we possess insufficient data. More importantly, the environments of the simulation need to be handled with caution as it may affect the results of the learning process.

With regard to the learning process, the distinguishing between anomalies and data outliers also require clean and sufficient data to improve its performance. The data must have been tested with various techniques such as closing time test, insulation resistance test, contact resistance test, etc. and approved by inspector, in order to allow the experts to make assumptions of the classification from these data. Moreover, the experts are always required for the parameters tuning process in every training process to achieve the practical model.

In order to choose input data for the Machine Learning process without involving feelings, the correlation matrix technique is used to determine the relation between each variable. However, if there is still a large amount of variables involved ; more consideration techniques may be required such as Principal Component Analysis (PCA). The PCA technique allows us to find the relation between variables in a more compact way, which leads to a smaller size of matrix. The smaller matrix leads to a less complex model, less model generated times, and easier to work with.

Lastly, after the model is generated ; it always needs to be tested by the experts to distinguish between anomalies and data outliers in order to improve the performance of equipment inspections. As a result, the electricity will be continuously generated with a lesser chance of a Force Outage incident that may affect the power stability in the electricity generating process.