BITS Pilani Pilani | Dubai | Goa | Hyderabad Work Integrated Learning Programmes	Introduction to Data Science M.Tech Data Science and Engineering [S2-20_DSECLZG523]	START DATE : 21-06-2021 END DATE : 22-08-2021

Anomaly detection

Introduction

Globally, Security of computers and the networks that connect them is increasingly becoming of great significance. Attacks on the nation's computer infrastructures are becoming an increasingly serious problem. Machine learning and data mining algorithms play important roles in designing intrusion detection systems. Misuse or signature detection techniques attempt to proactively detect the presence of such patterns so that any malicious attack on the infrastructure can be effectively defended against. In the anomaly detection approach, on the other hand, anomalous states in a system are identified based on a significant difference in the state transitions of the system from its normal states.

Objective

Follow the Data Science Methodology that we studied in Module 3. Analyze the data set, investigate and evaluate the result and predict the overall performance.

Dataset

The dataset is Anamoly.csv. It has been uploaded to canvas.

Tasks in this assignment

1. Write a Data Science Proposal for achieving the objective mentioned.
2. Perform exploratory analysis on the data and describe your understanding of the data.
3. Perform data wrangling / pre-processing (Module 8).
    a. E.g., missing data, normalization, discretization, etc.,
1. Apply any two feature selection engineering techniques (Module 8)
2. Compare the two selected feature engineering techniques.
3. Plot top 5, 6, and 8 features.
4. Provide a high-level description of Machine Learning models – association rules and random forest to predict.
5. Compare the performance of the two classifiers – association rules and random forest.
6. Present the conclusions/results in the format shared.

Expected Submissions

Two files are expected as the assignment submission.

1. The summary of the work in the template provided. (you may fill only the boxes relevant to this problem statement)
2. The executed ipynb file with clear subdivision of the codes and brief description of the purpose of respective code. All the executed tables or graphs and results should be present in the ipynb file. The ipynb file maybe submitted as a single .pdf file.