Project Description for Handover

Project Title: Hourly Forecasting of Message Volume and Anomaly Classification

Development Environment: Ubuntu

---

1. Hourly Forecasting of Message Volume (NMSG)

Objective: Develop a model using the NeuralProphet algorithm to forecast the hourly volume of incoming messages to the servers based on historical data.

Data Source: Sourced from the S3 production "Performance" bucket, stored in "/Data/train_data/".

Development Process:

- **Proof of Concept (POC):** Initiated in "notebooks/NeuralProphet_NMSG_testing.ipynb" before modularizing into the "src" folder.
- **Folder Structure:**
  - **"src/components":** Houses defined components utilized within specific pipelines.
  - **"src/pipelines":** Contains key pipelines for the project:
    - **1. ETL_pipeline.py:** For Export-Transform-Load operations on the data.
    - **2. Hyperparam_tuning.py:** Conducts hyperparameter tuning using MLflow for model optimization.
    - **3. Predict_pipeline.py:** Utilized for making predictions on the selected model post hyperparameter tuning.

Note: Uncommenting "-e ." in the initial run of requirements.txt activates setup.py, facilitating the use of Python modules from different locations. Following this, it can be commented out.

---

2. Anomaly Classification

Objective: Develop an anomaly detection model that utilizes the LSTM RNN algorithm to predict anomalies based on historical patterns, using the last 24 hours of performance data.

Development Process:

- **Proof of Concept (POC):** Currently in the POC stage, with development stages documented in the following notebooks:
  - **notebooks/LSTM_Classification_data_preparation.ipynb:** Details the data preprocessing stage.
  - **notebooks/LSTM_Classification_model.ipynb:** Focuses on the implementation of the LSTM model itself.
