**SLIIT Datathon 2024**

Challenge :

Solar power is one of the most efficient ways to generate clean and renewable energy. It helps reduce pollution and fight climate change. However, keeping solar power systems running efficiently isn’t always easy. Sometimes, things go wrong, and the energy output is not what it should be. These issues are called **anomalies**.

Anomalies can happen for different reasons, such as equipment problems, sudden weather changes and data errors.

Your role in this challenge is to act as an IT specialist monitoring the solar power system. You’ll carefully analyze IoT data from solar plants to find these anomalies. By identifying them, you help improve the way solar power systems work, making them more reliable and efficient.

Dataset(s) : [Link to dataset](https://mysliit-my.sharepoint.com/:f:/g/personal/datathon_sliit_lk/EjE-eggYKQZEqm-qHHvnEmgBzXEijvN3lvlF23_L_Waf-A?e=yBLsS1)

* Real-world data from solar power plants located at two(2) separate locations.
* 34 days of data captured at **15-minute intervals** for a detailed time-series analysis.
* Additionally, it includes weather data collected from IoT sensors at the corresponding locations.

Part A (Weight : 75%)

1. **Data Preparation:**

* Pre-process the data:
  + Prepare the data for analysis by cleaning and organizing it.
* Identify and fix issues in the dataset:
  + Identify and correct any issues in the data, such as missing values or inconsistencies.
  + Record each identified issue and describe the fix you applied.

1. **Exploratory Data Analysis (EDA):**

* Perform an in-depth analysis to understand the dataset and uncover insights.
* Leverage the **weather dataset** to provide additional context and identify potential anomalies in solar power generation.

*(Hint: Weather data can offer valuable clues for understanding anomalies.)*

1. **Anomaly Detection Model/Algorithm:**

* Develop a model/algorithm to detect anomalies in solar power generators.

*(Hint: Remember, you’re working with data points arranged over time. These patterns hold the power not only to predict the future but also to uncover hidden insights about the present.)*

* You can explore and implement one or more approaches, including:
  + Classical methods
  + Deep-learning-based methods such as LSTM, CNN-LSTM, DeepAR and Transformers.

1. **Results and Metrics:**

* Explain the steps to develop the anomaly detection model, including customizations made for the dataset and challenge requirements. Use text and screenshots to illustrate your approach.
* Evaluate your approach using metrics such as MAE, MSE, accuracy, and confusion matrix.
* Compare multiple models if time permits, and document their performance.

Part B (Weight : 20%)

1. **Store Output:**

* Load the dataset(s) and your detected anomalies (over time) into a database or a file you prefer (Ex: InfluxDB, PostgreSQL, MySQL, MongoDB, CSV, Excel, Parquet)
* Ensure the data is structured and accessible for visualization purposes.

1. **Power BI Dashboard:**

* Connect your database/DataSource to Power BI and create a interactive dashboard.
* Create a data model in Power BI with facts and dimensions for better organization and performance (optional task)
* Include screenshots of visualizations such as (but not limited to):
  + Power generation over time for each location/unit.
  + Irradiation vs. power generation.
  + Power generation vs. anomalies.

Part C (Weight : 5%)

* Research about Airflow and create a simple Airflow DAG containing tasks to pre-process, predict and load data.

You can use a self-hosted Airflow environment(docker/pip) or any managed Airflow service to test your DAG. You may submit UI screenshots and the DAG file for the final submission.

Submission :

* Document your work from the start in the report, including your thoughts, assumptions, and outputs (with screenshots).
* Submit the following items in your team’s designated folder:
  + Python Jupyter Notebook
  + Database Backup/Model output
  + Detailed Report
  + Power BI Desktop file (.PBIX) / Tableau Workbook
  + Airflow DAG file
* Submission Deadline: **Monday, December 23, 2024, by 4:00 PM.**
* You may optionally provide a 5-10 minutes video recording to elaborate on specific aspects and the dashboard, if desired

**Note**: To craft truly effective solutions, you need to think beyond what generative AI can offer. This hackathon is about showcasing your unique creativity and data-engineering skills.

You may use generative AI to research techniques and tools, but please refrain from using it for programming, as all code submissions will be reviewed for originality.

For any queries regarding the challenge, please contact via WhatsApp at **0720387828**.

- Happy coding, Good Luck !

**Additional Resources:**

* [Datasets](https://mysliit-my.sharepoint.com/:f:/g/personal/datathon_sliit_lk/EjE-eggYKQZEqm-qHHvnEmgBzXEijvN3lvlF23_L_Waf-A?e=yBLsS1)
* PowerBI desktop [download link](https://www.microsoft.com/en-us/download/details.aspx?id=58494)
* Tableau desktop [download link](https://www.tableau.com/products/desktop/download)
* Instructions to setup Airflow:
  + Using pip : [Medium link](https://www.freecodecamp.org/news/install-apache-airflow-on-windows-without-docker/)
  + Using docker : [Medium link](https://medium.com/@prithvijit.guha245/hello-world-airflow-docker-9102f4c5305b)