**IECDT 2024 Adaptation & infrastructure**

**Storm-Induced Nightlight Blackout Worldwide**

### **Overview**

Storms have a significant impact on critical infrastructure, often leading to widespread power outages. In this lab, you will explore the relationship between storm characteristics, socio-economic factors, and electricity blackouts, as identified through nightlight data. Drawing on real-world data from sources such as IBTrACS and Black Marble, along with a range of machine learning (ML) and data visualization techniques, you will clean, process, and develop predictive models to better understand the factors driving these outages.

### **Learning Outcomes**

By the end of this lab, you should be able to:

* Clean and preprocess storm and blackout datasets for analysis.
* Relate power outages to storm characteristics and socio-economic data.
* Apply and compare different ML techniques to predict power outages.
* Visualize results from predictive ML models.

### **TASKS**

#### **Task I: Clean IBTrACS Data for Storm Characteristics Input**

The IBTrACS (International Best Track Archive for Climate Stewardship) is a comprehensive dataset for tropical cyclones. You will clean and format this dataset for input into predictive models. The steps include:

* **Formatting**: Review the source file for inconsistencies or formatting errors. Ensure the data is in a consistent and analysable format.
* **Standardizing Records Across Agencies:** IBTrACS consolidates data from multiple agencies. Harmonize these variables to ensure uniformity, such as converting all wind speed measurements to the same standard.

#### **Task II: Get storm information for blackout events**

The blackout dataset, derived from sources like the Black Marble project, captures key details about power outages, including duration, affected customers, and geographic locations. To integrate storm-related information into this dataset, complete the following steps:

* **Matching records: match the IBTrACS data with the blackout data by a common identifier, i.e., storm ID.**
* **Extract storm information**: For each matched blackout event, extract relevant storm characteristics from the IBTrACS dataset, such as wind speed, pressure, and speed.

#### **Task III: Match Socio-Economic Data with Blackout Data**

Socio-economic factors such as population density, income, and urbanization levels are critical for understanding the context of power outages. To incorporate this information into the dataset, follow these steps:

* **Geographic matching**: Use location identifiers, i.e., latitude and longitude, to link socio-economic data with the blackout dataset.
* **Handle missing socio-economic data**: Identify missing socio-economic values in the matched dataset. Apply approximation techniques to fill in the empty data.

#### **Task IV: Predict Blackout Events Using Different ML Techniques**

To predict blackout events based on storm characteristics, blackout data, and socio-economic factors, follow these steps:

* **Train-test split**: Divide the dataset into training, testing, and evaluation subsets (e.g., 70%-20%-10% split).
* **Apply machine learning models**: Implement Extreme Gradient Boosting (XGBoost). XGBoost is effective for handling non-linear relationships and capturing feature importance. Train the model using the training set, with predictors including storm characteristics, socio-economic factors, and historical blackout data.
* **Hyperparameter tuning**: Optimize model performance by tuning hyper parameters such as: hyperparameters like learning rate and tree depth.
* **Model evaluation**: Evaluate the model's performance on the evaluation dataset using metrics like confusion matrix.

#### **Task V: Visualization model outputs**

Visualizing the results of your machine learning models provides deeper insights into performance. Potential outputs include analysis on the input variables, confusion **matrix** formodel performance, impact of hyperp**arameters on model development, and i**mportanceof the **predictors**

### **Key Questions to Explore**

* **What storm characteristics are most predictive of significant power outages?**
* **Which socio-economic factors contribute most to prolonged power outages?**
* **Does the timing of storm landfall affect outage duration?**

### **Deliverables**

1. **Cleaned Data**: Submit the cleaned and formatted blackout datasets along with the storm information and socio-economic data, i.e., the input for your model.
2. **Modeling Results**: Provide the trained machine learning models and a summary of their performance based on evaluation metrics.
3. **Three slides**: Summarize key insights, focusing on the storm characteristics and socio-economic factors that most influence power outages. Discuss any patterns or trends identified in the data.

### **TOOLS & TECHNOLOGIES**

1. **Python** (pandas, scikit-learn)
2. **Jupyter Notebook** for documenting and executing code

### **DATA SOURCES INFORMATION**

1. **IBTrACS (International Best Track Archive for Climate Stewardship)** <https://www.ncei.noaa.gov/products/international-best-track-archive>

This dataset provides comprehensive tropical cyclone data, including storm type, intensity, wind speed, and other storm characteristics.

1. **Black Marble (Nightlight Data)** <https://blackmarble.gsfc.nasa.gov/>

This dataset uses satellite observations of Earth’s night lights to capture changes in illumination due to power outages.

### **Conclusion**

In this lab, you will work through cleaning, processing, and analyzing a wide range of data related to storm-induced power outages. By applying machine learning techniques to this real-world data, you will gain practical insights into how storm characteristics and socio-economic factors contribute to global blackout events.