CHALLENGE ROUND *(Data Scientist)*

1. Problem/Assignment 1

Understanding the sun's position at any given time over a specific location is crucial for determining the shape and size of shadows cast by objects. Similarly, this principle can be applied to calculate the shaded area on a tilted surface that is inclined at a particular angle and orientation with respect to the North-South geographical line. In the foothills of the Himalayas, it is traditional to construct house roofs at a slope to facilitate efficient drainage of melting ice during winters. One such structure is a British Era Forest Bungalow in Chaukori, located in the Pithoragarh District of Uttarakhand (Latitude: 29.837508524171362, Longitude: 80.03159259597102). The Bungalow is oriented perfectly Southeast (45 degrees from due south towards the east) with a roof inclined at an angle of 30 degrees from the horizontal plane. The surrounding area is a large open lawn. At the center of the southeast-facing roof, there is a chimney that is 1 meter in height (measured from the higher bottom) with a horizontal cross-section of 1 square meter (a perfect square shape). You are tasked with developing an algorithm (preferably in Python) that can calculate the area of the shadow cast by the chimney on the tilted southeast-facing roof at any given time of day throughout the year, assuming clear sky conditions with no clouds or other obstructions.

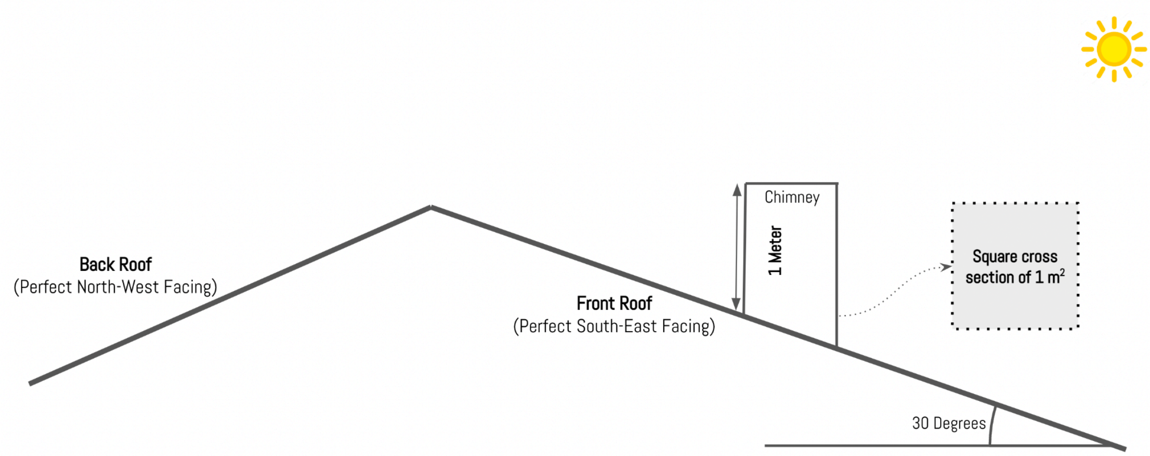
Required Submissions:

a. **Source Code:** Provide well-commented/documented source code of the algorithm. The code should take a specific date and time as inputs and output the area of the shadow casted on the souteast facing tilted roof in square meters.

b. **Solution Concept:** Submit a detailed explanation of your solution, including geometric and trigonometric diagrams, formulas, relationships, and calculations, compiled in a PDF document.

c. **Example Calculation**: Calculate and provide the area of the shadow cast by the chimney on the tilted southeast-facing roof on October 8, 2022, at 11:00 AM IST.

d. **Smallest Shadow:** Identify and provide the date and time when the chimney casts the smallest shadow on the tilted roof during the calendar year 2022.



Figue 1: Explains the geometrical setup of the roof and chimney

2. ***Problem/Assignment 2***

Photovoltaic (PV) power systems are the fastest-growing renewable energy technology globally. They are extensively deployed to meet rising energy demands and mitigate climate change. A major advantage of PV systems is their modular design: PV modules connect in series to form a string, and multiple strings connect in parallel to create a PV array. This array then connects to an inverter, which includes a maximum power point (MPP) tracking system for optimal efficiency. Expanding system capacity is easy—just add more panels. Modern solar inverters also feature SCADA systems that log current (I_{mpp}) and voltage (V_{mpp}) at regular intervals. Analyzing this data with advanced analytics or machine learning algorithms provides valuable insights.

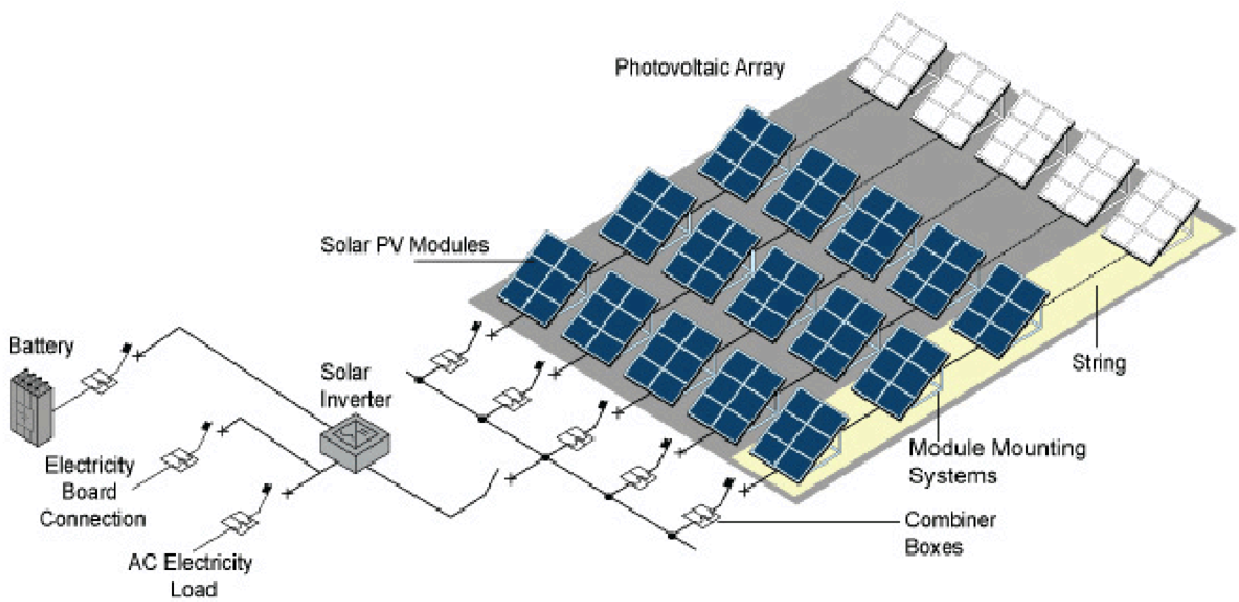


Figure 2. Basic design of Photovoltaic power system

In normal conditions, a PV system should perform as expected and generate power ($P_{Expected}$) based on the installed capacity, incident solar irradiation, and module temperature. The formula for calculating instantaneous expected power is given below (for your convenience, values are already calculated and added in the shared data).

$$I_{expected} = I_{rated} \bullet \left(\frac{G}{1000}\right) \bullet [1 + \alpha(T_{module} - 25)]$$
$$V_{expected} = V_{rated} \bullet \left(1 + mT_{module} + b\right) \bullet LN\left(\frac{G}{1000}\right) \bullet [1 + \beta(T_{module} - 25)]$$
$$P_{expected} = V_{expected} \bullet I_{expected}$$

Where, I_{rated} and V_{rated} are current and voltage values at maximum power point under STC conditions as mentioned on nameplate of PV modules. The α (0.06% per degree change in T_{module}) and β (0.27% per degree change in T_{module}) are temperature coefficients of current and voltage respectively. T_{module} is module temperature. G is instantaneous irradiance. The m (=2.4/1000), and b (=6.02/100) are constants specific to c-Si PV module technology.

>>>** Any anomaly in the system's operating conditions can lead to a significant deviation in power production from the expected production profile. Anomalies can belong to different types (classes) due to various reasons (e.g., shading, increasing dust layers on the PV module, defects in PV modules, inverter or power electronics failure, etc.)**. The following two conditions can be used to detect such anomalies in the time-series data:

a. *If there is a significant difference between the rate of rise/decline of expected power and the actual power. In normal situations, the expected and actual power shall follow a similar trend (simultaneously rise or decline according to the irradiation and module temperature).*
b. *If the difference between the expected and actual power increases rapidly and exceeds the expected deviation threshold. During faulty conditions, the actual power yield of the system is significantly low compared to the expected power yield.*

We have prepared sample data of almost a month that is logged at a time granularity of 1 minute. The sample data belongs to 1 MPPT and contains information on the following parameters:

1. Date-time stamp (Datetime)
2. Actual MPPT Power (P)
3. Expected MPPT Power (P_exp)

>>>> Can you deploy deep learning or machine learning-based techniques to build a model using the shared data and detect the anomalies in the system? You can summarize the model's performance by sharing the following findings:

1. What do you think about the data distribution? How are you defining the data point as faulty or Normal? Which model will you develop for this particular problem, and why?
2. How did you train the models to understand the normal and abnormal data points (or instances) without explicit labeling? You can share your source code with nicely commented reasoning/assumptions.
3. Run the model on the given data set to identify the anomalies. You share some visualizations with the explicit demarcation of anomalous and normal instances in the data set? Reasoning or logical thinking is acceptable.

Note: The data file is attached separately.