***Objective***

The variable to be predicted is continuous (energy production). Therefore, this is an Regression project. The primary goal is to model energy production as a function of environmental variables

***Data Set Description***

The file solarpowergeneration.csv contains the data for this example. Here, the number of variables (columns) is 10, and the number of instances (rows) is 2920. We have the following variables for this analysis:

* **distance\_to\_solar\_noon**, in radians.
* **temperature**, daily average temperature, in degrees Celsius.
* **wind\_direction**, daily average wind direction, in degrees (0-360).
* **wind\_speed**, daily average wind speed, in meters per second.
* **sky\_cover**, on a five-step scale from 0 to 4, with 0 completely clear and 4 wholly covered.
* **visibility**, in kilometers.
* **humidity**, in percentage.
* **average\_wind\_speed**, average wind speed during the 3 hours de measure was taken in meters per second.
* **average\_pressure,** average barometric pressure during the 3 hours the measure was taken in, in mercury inches.
* **power\_generated(Target)**, in Jules for each 3 hours

|  |  |  |
| --- | --- | --- |
| **Milestone** | **Duration** | **Task start - End Date** |
| Kick off Business Objective discussion | 1 day | 01-Apr-2025 |
| Dataset Discussion | 2 days | 03-Apr-2025 |
| EDA | 1 Week | 08-Apr-2025 |
| Model Building | 1 Week | 15-Apr-2025 |
| Deployment |  | 22-Apr-2025 |
| Final presentation | 1 Day | 29-Apr-2025 |

Protocols:

1. All participants should adhere to agreed timelines and timelines will not be extended.
2. All the documentation – Final presentation and python code to be submitted before the final presentation day.
3. All the participants must attend review meetings.