Introduction

Accurate solar irradiance forecasting is a cornerstone in advancing photovoltaic (PV) energy systems, which play a central role in the global transition to clean and sustainable energy. As the EU intensifies efforts to meet its climate targets, the expansion of PV installations—both large-scale and residential—has become increasingly vital. However, the intermittency of solar energy, driven by local meteorological conditions and diurnal cycles, poses significant challenges for reliable integration into the power grid. The variability and unpredictability of PV production not only complicate reserve capacity planning but also hinder the development of decentralized energy systems, such as energy communities, which rely on stable and predictable supply-demand dynamics.

This mini-contest focuses on a regression task where the goal is to predict **solar irradiance** for a specific location using a set of **meteorological** and **physical variables**. Since irradiance directly correlates with PV power output in a nearly one-to-one relationship, improving irradiance predictions effectively translates into improved production forecasts. Such predictive capabilities are essential for enabling grid operators and end users to perform more effective yield planning, storage management, and electricity market participation. By promoting more reliable and trustworthy forecasts, this task aims to contribute to reducing generation-load mismatches, minimizing energy waste, and fostering the economic viability of local energy trading models.

Ultimately, the challenge offered in this mini-contest provides an opportunity to explore how ML models can capture the complex, non-linear relationships between environmental variables and solar irradiance. You will gain hands-on experience with a problem of real-world relevance and contribute to the broader goal of supporting the widespread adoption of photovoltaic energy.

Features' Description

Meteorological data was collected for free via the Open Meteo API, a commercial provider, by inputting the geographical coordinates of the PV plant.

Intermediate irradiance data was generated using the clearsky model provided by pvlib, a free and open-source Python library for photovoltaic system modeling.

Actual Irradiance data—the target variable—was directly measured on-site by a pyranometer installed on the plane of array.

Variable	Unit	Description
temperature_2m	°C	Air temperature at 2 meters above ground
relative_humidity_2m	%	Relative humidity at 2 meters above ground
dew_point_2m	°C	Dew point temperature at 2 meters above ground
apparent_temperature	°C	Apparent temperature is the perceived feels-like temperature combining wind chill factor, relative humidity and solar irradiance
precipitation	mm	Total precipitation (rain, showers, snow) as sum of the preceding hour.
rain	mm	Only liquid precipitation of the preceding hour including local showers and rain from large scale systems.
snowfall	cm	Snowfall amount of the preceding hour in cm.
pressure_msl surface_pressure	hPa	Atmospheric air pressure reduced to mean sea level or pressure at surface.
cloud_cover	%	Total cloud cover as an area fraction
cloud_cover_low	%	Low level clouds and fog up to 2 km altitude
cloud_cover_mid	%	Mid level clouds from 2 to 6 km altitude
cloud_cover_high	%	High level clouds from 6 km altitude
et0_fao_evapotranspiration	mm	ET_0 Reference Evapotranspiration of a well watered grass field. ET_0 is calculated from temperature, wind speed, humidity and solar irradiance.
vapour_pressure_deficit	kPa	Vapor Pressure Deificit (VPD) in kilopascal (kPa). For high VPD (>1.6), water transpiration of plants increases. For low VPD (<0.4), transpiration decreases
wind_speed_10m wind_speed_100m	km/h km/h	Wind speed at 10 or 100 meters above ground. Wind speed on 10 meters is the standard level.
wind_direction_10m wind_direction_100m	°	Wind direction at 10 or 100 meters above ground
wind_gusts_10m	km/h	Gusts at 10 meters above ground of the indicated hour. Wind gusts in CERRA are defined as the maximum wind gusts of the preceding hour.
soil_temperature_0_to_7cm soil_temperature_7_to_28cm soil_temperature_28_to_100cm soil_temperature_100_to_255cm	°C	Average temperature of different soil levels below ground.
soil_moisture_0_to_7cm soil_moisture_7_to_28cm soil_moisture_28_to_100cm soil_moisture_100_to_255cm	m^3/m^3	Average soil water content as volumetric mixing ratio at 0-7, 7-28, 28-100 and 100-255 cm depths.
is_day	/	0 = night; 1 = day
apparenth_zenith	°	Apparent sun zenith accounting for atmospheric refraction.

zenith	°	Solar zenith angle
apparent_elevation	°	Apparent sun elevation accounting for atmospheric refraction.
elevation	°	Actual elevation (not accounting for refraction) of the sun in decimal degrees, 0 = on horizon. The complement of the zenith angle.
azimuth	°	Azimuth of the sun in decimal degrees East of North. This is the complement of the apparent zenith angle.
equation_of_time	min	Difference in time between solar time and mean solar time.
ghi	W/m^2	Global horizontal irradiance
dni	W/m^2	Direct normal irradiance
dhi	W/m^2	Diffuse horizontal irradiance
tracker_theta	°	Tracker rotation angle as a right-handed rotation around the axis
aoi	°	Angle of incidence of the solar vector on a surface. This is the angle between the solar vector and the surface normal
surface_azimuth	°	Azimuth angle of the surface
surface_tilt	°	Panel tilt from horizontal. For example, a surface facing up = 0°, surface facing horizon = 90°
POAb	W/m^2	Component of direct sunlight that is incident on the surface of a PV module
POAg	W/m^2	Component of direct sunlight that is reflected off the ground and then strikes on the surface of a PV module
POAd	W/m^2	Component of direct sunlight that strikes on the surface of a PV module after having been scattered by molecules or particulates in the atmosphere.
Irr0	W/m^2	Modelled irradiance
Irr (target)	W/m^2	Total irradiance received on the PV plane of array as average of the preceding hour.