Prediction of Parking Area Solar Panel Electricity Generation at the University of Liège

A. Birtles, G. Delporte, R. Lambermont & A. Louis

Université de Liège

May 2024

Table of Contents

Introduction

We aim to predict the daily production of solar panels of two parking at the University of Liège.

Context

- The objective is to obtain solar panels predictions for the upcoming days to facilitate scheduling purpose (cost, savings, maintenance, ...).
- The users of our project are the administrators at the University in charge of the solar panels.
- Challenges reside in the lack of awareness regarding electricity production and the absence of a connection with the weather station.

The concept is to incorporate electricity production forecasts into a user-friendly dashboard, streamlining user experience and facilitating convenient access to solar panel maintenance and usage.

ML Canvas

THE MACHINE LEARNING CANVAS

Designed for: MLOPS Triple P Designed by: A. Birtles, G. Delporte, R. Lambermont, A. Louis Date: 05/01/2024

PREDICTION TASK

Type of task? Regression (predicting a continuous value - the amount of electricity generated).

Entity on which predictions are made? Photovoltaic (solar) panels.

Possible outcomes? Predicted watt-hours of electricity produced by the panels.

Wait time before observation? Immediate for predictions; actual observation wait depends on the data collection intervals (e.g., every 15 minutes).

DECISIONS

How are predictions turned into proposed value for the end-user? Predictions can be integrated into a Google Calendar, allowing the administrator to visualize and manage energy production schedules, maintenance, and usage optimization. This aids in planning and operational efficiency without manual inference.

VALUE PROPOSITION

Product: Friendly user calendar with the prediction of electricity.

Alleviates: The users will just have to look at the prediction and not have to infer themselves about the production.

Advantages: Better agreement between teams inside Uliege and easier access and/or use of the photovoltaic panel maintenance and usage

DATA COLLECTION

Strategy for initial train set & continuous update: Utilize historical data from the University's dataset.

Trying to get new data with the help of the University.

DATA SOURCES

Where can we get (raw) information on entities and observed outcomes? Data will be sourced from the University's private dataset. Additional data may be obtained from integrated meteorological sensors and APIs that provide real-time weather data, from the weather station of the University.

IMPACT SIMULATION

Can models be deployed? Yes, models can be deployed after testing.

Cost/gain values for (in)correct decisions? Cost includes potential inaccuracies in energy management and maintenance scheduling; gains from accurate predictions include cost savings, optimal resource utilization, and improved maintenance scheduling.

Fairness constraint? Ensure the model's predictions do not inadvertently favor certain times/days.

MAKING PREDICTIONS

Time available for this + featurization + post-processing? Immediate processing required; limited to a few seconds to ensure timely updates.

Compute target? Must be lightweight enough to run on existing university servers or cloud infrastructure without significant cost increases.

BUILDING MODELS

Techniques: Start with basic models (Linear Regression, Decision Trees), then explore ensemble methods) and a small neural network for better performance as well as a neural network that outputs some distribution parameters..

FEATURES

Input representations available at prediction time, extracted from raw data sources: Features might include time of day, historical energy production data, current and forecasted weather conditions. WE want to have some sort of dashboard that will display the results for decision making.

MONITORING

Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)? Key performance indicators like Mean Squared Error (MSE)

OWNML.CO

Data

The data used in this project consists of two main parts:

- Solar power production, the target variable
- Forecast data from the university's laboratory of climatology, explonatory variables for the temperature, irradiance, humidity, ...

At the beginning of the project we made some simple EDA before creating our models to try predict the power production of the different pannels using the explonatory variables

Models

Two approaches were undertaken, simple scikit-learn models (LR, KNN, DT, RF, GBM, SVM) and pytorch neural networks. For the training of those models, we used a 75%/12.5%/12.5% train/test/validation split.

We quickly found out that the random forests were our best possible choice. We then used a randomized grid search to look for better parameters.

We also created a Weight & Biases instance to check in real time if the training of our models was going well.

W&B

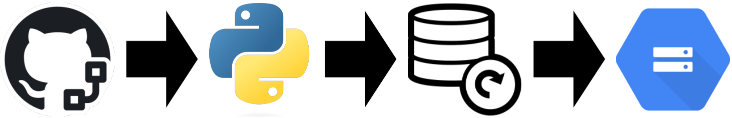
Tools Used - Collaboration

Throughout the project we used diverse tools to make the collaboration between members easier:

- **Code Sharing**: We used GitFlow and GitHub to collaborate and exchange the different pieces of code created by the different group members.
- **Planning**: We used a Trello board to organize and plan the different tasks that needed to be done
- **Communication**: We used Discord as the main point of communication for the project, taking advantage of the writing and talking channels.

Tools Used - Cloud Computing Automation

The maintenance of the models on the cloud is automatic; a GitHub Action is triggered every day at 2:30 AM, launching `.py` scripts, updating the data used for predictions and the models, and sending the computed data to a Google Cloud Bucket.

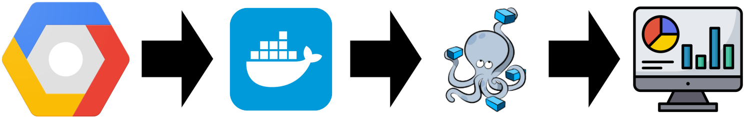


Application - New Data

LIÈGE université
Climatologie & Topoclimatologie

- **Training Data:** Data collected in 2022 from the meteorological institute of the University of Liège, with readings every 15 minutes.
- **Future Data Access:** Attempts to access future data from the meteorological institute were unsuccessful, hindering real-time prediction of solar panel electricity production.
- **Modified Data Approach:** Modified the 2022 dataset by computing means, standard deviations, and adding random noise to simulate real-world conditions.
- **Data Bounds:** Modified data adhered to predefined bounds, ensuring it remained within realistic ranges for features such as cloud cover, wind speed, and humidity.

Application - Dashboard



- **Docker Image:** We create a Docker image to set up our prediction visualization environment.
- **Deployment with Docker:** Using Docker and Google Cloud Platform, we deploy our environment to the cloud.
- **Accessing Computed Data:** We access pre-computed prediction results stored in a Google Cloud Bucket.
- **Showing Predictions:** With this setup, we showcase the predictions derived from the accessed data, with different visualization possibilities (power, irradiance, ...) for a different number of days.

Application - Dashboard (cont'd)

SCAN ME

Final Touches and Conclusion

Final touches:

- Commenting the code
- Improve the project's repository documentation
- Linting the code

With this first project in the field of MLOps now behind us, we can state that we had to start over again, we would probably continue to follow the framework seen during the course for later projects. The project would also have been more interesting if we could have gotten our hands on the actual data of the climatology laboratory. We could thus have trained our model on various types of predictions to make it more robust (e.g. 1 to 15 days predictions) with error margins to make it interesting to use to maintain the solar power plant.