solar data processing using statistics and ML
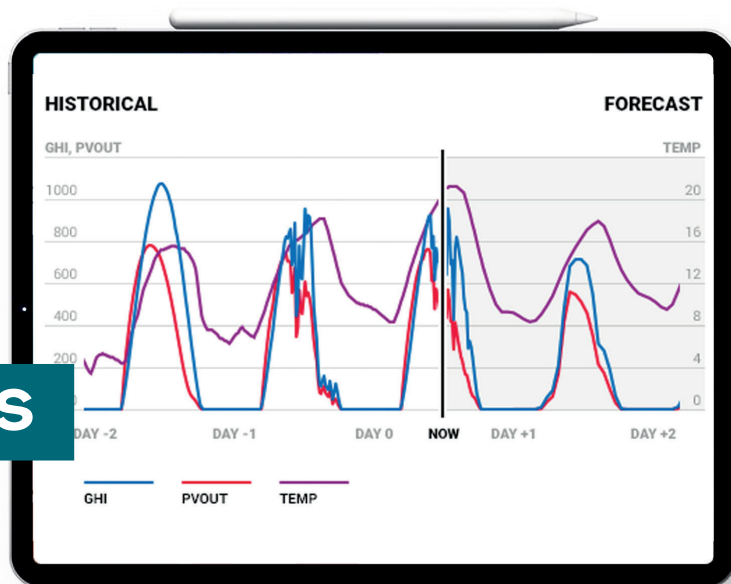
Solargis provides reliable and accurate solar, weather and solar electricity data that are used in the whole lifecycle of solar power plants, from prospection to development and operation. Since 2010 Solargis develops and operates a platform for fast access to historical, recent, and forecast data for almost any location on the Earth.

about project

The first objective of the project is to automatize the quality control of various data related to the solar energy sector. The second part of the project is dedicated to fault detection of solar photovoltaic power plants. First, the problem is explored using statistical analysis. The outcomes of statistical analysis are then used to build and improve statistical models and simple machine learning models. Furthermore, the results are attempted to be improved using deep learning, such as using the outputs of statistical models as labels in classification tasks.

client

facts & numbers

7 areas - found to be candidates for automation and optimization

Result – manual work reduction of data operators from hours to minutes

technology stack

- Docker
- Python
- Pandas, NumPy, SciPy
- Matplotlib
- Scikit-learn
- Keras, TensorFlow

To have a scalable and comprehensive solution, in collaboration with the team of client, we helped to extend the data infrastructure and processes for this project. The infrastructure includes preparing dedicated databases, algorithmic modularization and ends with automated testing and reporting. From the process perspective, the team is working in an agile environment

challenges

Create a robust solution which suits the given problem the best.

By tackling the problem from both statistical and machine learning perspective we can propose comprehensive solutions covering a large spectrum of aspects. Moreover, the combination of our scientific and engineering expertise enables us to deeply understand the problems and define a region of validity of the given solution.