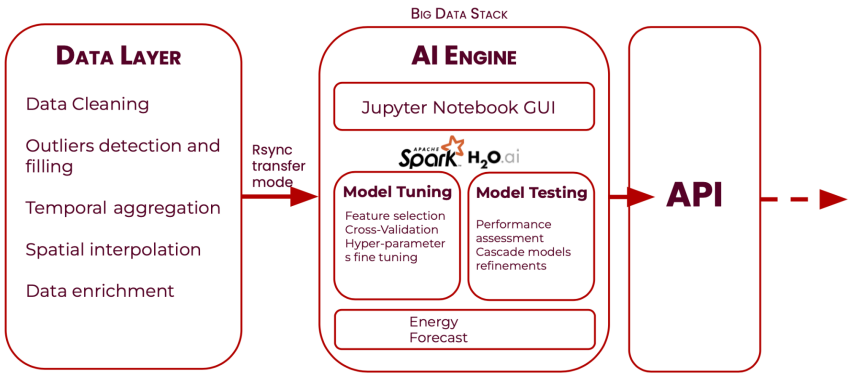
Technical Specification Double-side Page

NAME OF THE COMPANY: AMIGO SRL
CHALLENGE: REACH-2022-THEMEDRIVEN-CERTH_4.1 - RES FORECASTING
TRACK: THEME DRIVEN

1. TECHNICAL SCOPE:

The **PENGUIN** (AI-Powered rEnewable eNerGy soUrces forecastINg) solution addresses the **RES Forecasting Theme challenge** with a novel Deep Learning (DL) pipeline tailored for **photovoltaic (PV) energy generation forecasting**. Trained, tested, and validated on PV data from the SmartHome prototype provided by **CERTH**, PENGUIN offers reliable **data-driven day-ahead hourly predictions**. PENGUIN's architecture encompasses a data layer performing data integration and data preprocessing operations, an AI engine, and an API that delivers the PV forecasts to the user.
The AI engine is the core of PENGUIN, and comprises **a set of season-specific, multi-layer Machine Learning** (ML) **model**s receiving ready-to-use features from the data layer. Features include the characteristics of the PV plant and environmental variables driving PV energy generation—like short-term weather forecasts from the COSMO model and atmospheric and climate forecasts from CAMS (e.g. total dust, PM2.5, solar radiation).
PENGUIN's day-ahead 15-minutes PV forecasts achieve an average **nMAPE of 17.7%**, providing predictions with a temporal resolution crucial for energy management and dispatching. Such predictions are released on a daily basis and are made available through an API service, which is the final layer of the MVP.

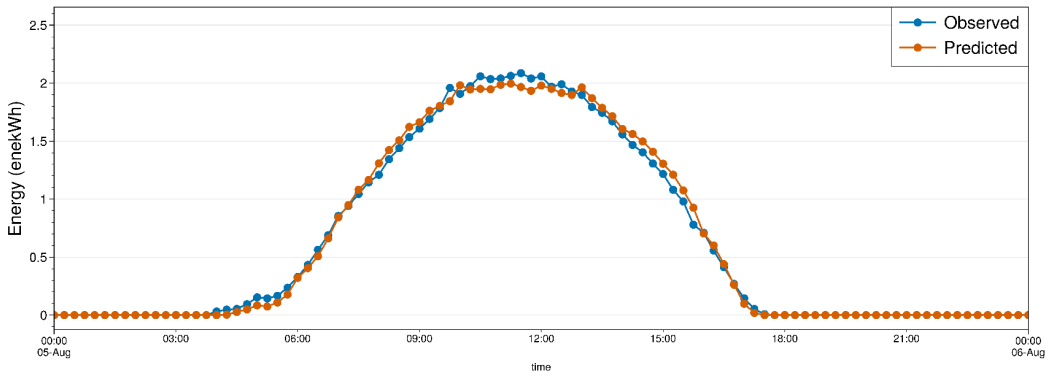


2. ALGORITHMS, TOOLS AND CONCLUSIONS:

The **PENGUIN AI engine** operates using a **three-model cascade pipeline**. This pipeline first processes daily data, then, converts it to hourly temporal resolution and then to 15-minute intervals. Predictions from the daily model, followed by those from the hourly model, are forwarded to the subsequent model with finer time resolution as additional features. The models chosen belong to the feed-forward multilayer artificial neural network family.
For each of the three models composing the cascade, a hyperparameter optimization phase is undertaken using a GridSearch strategy with Cross Validation. This phase is performed once and is aimed at finding out the optimal configuration that minimises error while maintaining computational efficiency.
The entire AI structure is **integrated with the Big Data Stack provided by ITI Datahub**. This integration facilitates cluster execution in a Spark environment using the PySparkling AI engine. Our findings indicate that the most crucial features are temperature, wind components, and total cloud cover from COSMO weather forecasts, and the angle of irradiance between the perpendicular to the surface of the PV modules and solar radiation. For every weather variable among the selected features, data from the four nearest grid-points are taken into account. The networks resulting from the optimization phase use include three neuron layers with a Maxout activation function. They require between 500 and 700 epochs to achieve optimal outcomes. To enhance accuracy by accounting for seasonal patterns in weather variables and solar irradiance, a dedicated three-model cascade pipeline is trained for each season.

Following discussions with the Data Provider, the **nRMSE** is selected as a metric to assess model performance. The accompanying figure illustrates the results of a day-ahead forecast example compared with observed values at 15-minute intervals. The table presents the nRMSE for each season, calculated over a one-month validation period. Regardless of season, PENGUIN's results are **compliant with the goal of the challenge**, which is to achieve an error rate of less than 20% for day-ahead forecasting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951981



	Winter	Spring	Summer	Autumn
nRMSE	18,3%	19,9%	16,9%	15,8%

3. SCALABILITY AND FLEXIBILITY OF THE SOLUTION:

The **PENGUIN solution is versatile** and can be readily adapted to any PV generation system with minimal adjustments to new input data. Importantly, its Deep Learning models rely solely on features derived from commercially available open data. We have planned a **dual-execution strategy**: a) real historical data from the PV generation system is used for training, allowing for a customised forecasting service; b) synthetic PV production data, derived from geographical location and system characteristics, is used for training.
This approach guarantees robust scalability. It means the service can be **tailored to diverse customer requirements**, offering flexibility in business strategy. Since, at this stage, only solution a) has been implemented, the product is estimated to be at a **TRL of 5/6**.
Furthermore, we are currently **discussing with OmegaLambdaTech**, who is also part of the REACH incubator, about integrating their energy consumption forecast data from ecostar3 into the PENGUIN product. This collaboration is aimed at fostering a multi-stakeholder data value chain.

4. DATA GOVERNANCE AND LEGAL COMPLIANCE:

The PENGUIN solution offers a service that delivers daily forecast data in the form of an **API**. This API prioritises security, incorporating **strong authentication, authorization procedures, and data encryption protocols**. We utilise OAuth 2.0 for authentication, enabling external users to safely access the product without directly sharing their credentials. Instead, users grant limited access to their protected resources. The security infrastructure is built on AWS services including Cognito, API Gateway, and Lambda, ensuring user authentication through private tokens.

In compliance with the General Data Protection Regulation (**GDPR**), all data processed by our software is anonymized and aggregated to prevent identification of individual consumers, thereby respecting privacy and eliminating any risks related to personal data handling. In the event of a data breach, we have a robust incident response plan to identify and mitigate the impact swiftly. We also follow the principle of data minimization, only collecting and processing data that is absolutely necessary for the provided services. We conduct regular data protection impact assessments to identify and reduce data protection risks, ensuring our compliance with GDPR.

5. QUALITY ASSURANCE AND RISK MANAGEMENT:

We followed covered QA for three different aspects:
DATA QA: We performed a data cleaning procedure on the PV data, provided by CERTH. In particular, we validate incoming data to ensure that it matches specified criteria, such as value ranges for meteorological values and geographic position. Our data cleaning process checks the data and identifies and corrects discrepancies, missing values and outliers.
CODE QA: For each feature or function that is developed, appropriate tests are written before the actual code is created. This ensures that all code is testable and tested.
OUTCOME QA: To assess the accuracy of our predictions, we compare them retrospectively with actual power generation and adjust our models as necessary based on these analyses. We use performance metrics such as normalised root mean square error (nRMSE) to quantify the quality of our forecasts.no
Lastly, we have identified product scalability as potential risk. To mitigate this, we have established the dual-execution strategy outlined in paragraph 3.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951981

Means for accessing the MVP

To execute an API request of the next day forecast, first authenticate following these steps:

- Navigate to https://penguin-reach.auth.us-east-2.amazoncognito.com/login?client_id=3jhosh3bo09bhktqne69ck07s2&response_type=code&scope=aws.cognito.signin.user.admin+openid&redirect_uri=https%3A%2F%2Fhhnyizrky4.execute-api.us-east-2.amazonaws.com%2Fpenguin%2Foauth2callback
- Type
  - username: penguin-reach-demo
  - password: Bilbao@2023
- Copy the ID TOKEN provided, it is mandatory to authorise the request

Connect to the API endpoint https://hhnyizrky4.execute-api.us-east-2.amazonaws.com/penguin/ with the preferred method, specifying the bearer token inside the header.

API documentation is available here: https://app.swaggerhub.com/apis/amigoapitest/Penguin-Demo/1.0

Example Usage by using Postman:

- Paste the API address in the URL bar.
- Select GET request and specify to */data* endpoint at the end of the API address.
- Add *date* parameter, in the format YYYY-MM-DD (only April 2023 is now available as test dataset)
- Add the token in the Authorization tab, choosing Bearer Token from the menu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951981