IANOS

SUSTAINABLE SOLUTIONS

for islands' decarbonisation

iVPP Forecasting Engine v1.0

AUTHORS:

P. Tzallas, N. Bezas, I. Moschos, D. Ioannidis [CERTH]

H2020-LC-SC3-2018-2019-2020 / H2020-LC-SC3-2020-EC-ES-SCC

EUROPEAN COMMISSION

Innovation and Networks Executive Agency

Grant agreement no. 957810

PROJECT CONTRACTUAL DETAILS

Project title	IntegrAted SolutioNs for the DecarbOnization and Smartification of Islands
Project acronym	IANOS
Grant agreement no.	957810
Project start date	01-10-2020
Project end date	30-09-2024
Duration	48 months
Project Coordinator	João Gonçalo Maciel (EDP) - JoaoGoncalo.Maciel@edp.com

DOCUMENT DETAILS

Deliverable no.	D4.3
Dissemination level	Public
Work package	WP4 IANOS Multi-Layer VPP Operational Framework
Task	T4.2 Grey-box modeling & AI meta-learning for accurate energy demand / generation estimation
Due date	31st March 2022
Actual submission date	31st March 2022
Lead beneficiary	CERTH

Version History

v	Date	Beneficiary	Changes
0.1	06/09/2021	CERTH	TOC
0.2	11/10/2021	CERTH	Literature Review and methodology
0.3	13/01/2021	CERTH	Methodology implementation
0.6	21/02/2022	CERTH	Consolidated version for review
0.7	14/03/2022	ETRA I+D	Peer revision

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

0.8	28/03/2022	ENG	Peer revision
1.0	29/03/2022	CERTH	Final Version

Disclaimer

This publication reflects the authors' view only and the European Commission is not responsible for any sue that may be made of the information it contains.

Table of contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

List of Tables

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

List of Figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

1 Document Information

1.1 Executive summary

This document is part of the deliverables of the IANOS project, in the context of Work Package 4 (WP4) “IANOS Multi-Layer VPP Operational Framework”. Specifically, it presents the first version of IANOS iVPP forecasting engine, which is responsible for providing the necessary forecasts for all the uncertainties in every time horizon and spatial distribution. The deliverable provides information about the different sub-components of IANOS iVPP Forecasting Engine, namely load, PV and Wind generation and energy market forecasting.

Initially, the introductory sections provide the necessary information about the deliverable. The scope and the objectives of the deliverable is presented and, subsequently, the structure of the current deliverable as well as its relation to other tasks.

Before the presentation of IANOS forecasting engine, a valuable initial literature review of the related works in energy related time series forecasting is held. In the last decades the attention of the researchers has been focused in problems related to energy forecasting, due the increasing inclusion of renewable energy sources, and their uncertainty that introduces fluctuations to the energy grid. Alongside the literature review, the different energy markets in Europe are analysed, highlighting the importance of future predictions of its values.

Subsequently, the allocation of the forecasting engine component into the IANOS project’s overall architecture is presented. The interconnections of the different components to the module are presented together with the outer dependencies. Furthermore, an initial daft of the data model utilized by IANOS iVPP Forecasting Engine is provided, in addition to the different software pre-requirements of the module.

In the core of the document, the methodology followed within the developing of the forecasting engine is presented. The initial step of IANOS forecasting methodology is the process of data collection, which corresponds to the connection to IANOS ESB module. The next step is the pre-processing of the available data, and the detection of the existing outliers. IANOS iVPP Forecasting Engine is separated into load, PV generation, wind generation and energy market forecasting sub-modules. Based on the aforementioned categorization the document presents an exploration of the available data for each category, the strategy utilized alongside with the corresponding features, and finally the predictions models that are used to accumulate accurate forecasts, which vary from tree based models to support vector machine models and artificial neural network models.

After analysing the methodology, there is the presentation of the metrics used in the evaluation of the results, which are the MAE, RMSE, MAPE, sMAPE and MMR metrics. Subsequently there is the demonstration of actual results. For each submodule a table is presented with the collective results of each model used, and an example plot of the best performing algorithm.

In the last section of the deliverable there is the conclusions and findings of the analysis of IANOS iVPP forecasting engine, with the next steps and future work of the task, concerning the second version of the deliverable.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 957810

1.2 List of acronyms and abbreviations

Abbreviations	Full Description
ANN	Artificial Neural Network
ARIMA	Auto Regressive Integrated Moving Average
CNN	Convolutional Neural Network
GBR	Gradient-Boosted Regressor
GRU	Gated Recurrent Units
LightGBM	Light Gradient Boosting Machine
ENTSO-E	European Network of Transmission System Operators for Electricity
FCR	Frequency Containment Reserve
IoT	Internet of Things
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Errors
ML	Machine Learning
MLP	Multilayer Perceptron
MMR	Mean/Mad Ratio
NL	Netherlands
PV	Photovoltaic
RBF	Radial Basis Function
RF	Random Forest
RNN	Recurrent Neural Network
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
sMAPE	symmetric Mean Absolute Percentage Error
SVM	Support Vector Machine
SVR	Support Vector Regressions
VPP	Virtual Power Plant
RBFNN	Radial Basis Function Neural Network
kNN	k-Nearest Neighbors
ENN	Elman Neural Network
GRNN	Generalized Regression Neural Network
EV	Electric Vehicle
ESB	Enterprise Service Bus
MAD	Mean Absolute Deviation
ANFIS	Adaptive Network based Fuzzy Inference System
NARXNN	Nonlinear Autoregressive Neural Network with Exogenous Inputs
ELM	Extreme Learning Machine

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

SCRM	Clear Sky Solar Radiation

2 Introduction

2.1 Scope and objectives of the deliverable

The current deliverable provides information about the methodology that was followed in defining and developing the necessary models for the IANOS iVPP Forecasting Engine. Additionally, a thorough literature review of the previous works in energy forecasting, contacted in the context of the project, is presented.

The deliverable presents different predictive tools, namely for demand or consumption forecasting, for generation forecasting, for both wind and PV generation, and for energy market forecasting considering different time frames. Each tool utilizes analytical, or data driven techniques derived from the current State-of-Art methods and further expanding them. Finally, each tool is evaluated with the appropriate evaluation metrics, and tuned to achieve the best possible forecasting results.

The forecasting tools described in the deliverable constitute the core of iVPP Forecasting Engine, developed within the IANOS project iVPP framework.

2.2 Structure of the deliverable

The structure of the deliverable is as follows:

- **Chapter 1** presents information about the current document, namely the executive summary of the deliverable and a list of acronyms and abbreviations.
- **Chapter 2** presents the scope, objectives and structure of the deliverable, as well as its relation to the other tasks and deliverables of the IANOS project.
- **Chapter 3** provides the initial literature review for relative works, which contains information for load forecasting, PV and wind generation forecasting, and energy market forecasting.
- **Chapter 4** provides information about the architecture of iVPP Forecasting Engine, within IANOS project. Additionally, the model models utilized are described, together with the software pre-requirements.
- **Chapter 5** is the core of the document, which describes the methodology that was employed to develop the forecasting tools. Initially, the data collection step is described, which corresponds to the connection to ESB module of IANOS project. Subsequently, the preprocess step is described, followed by the description of the different forecasting tools.
- **Chapter 6** describes the metrics used for the evaluation of the forecasting algorithms and the evaluation results.
- **Chapter 7** provides conclusions about the overall work implemented in the current deliverable.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

2.3 Relation to other tasks and deliverables

This deliverable draws input from the defined Use Cases and requirements in D2.1 “Report on Islands requirements engineering and UCs definitions”. Additionally, in T2.5 “System Architecture”, which provides a description of the system’s architecture, there is an initial description of the functionality of the component described in this deliverable, and its interconnections and dependencies.

The current deliverable is closely related to all the tasks from WP4. More specifically, in T4.1 “Cyber-Secure data monitoring and VPP Governance”, the Enterprise Service Bus (ESB) is described, which is responsible to provide the Forecasting Engine with the necessary input data. Moreover, the component uses inputs from T4.3 “Intelligent VPP Clusters' Segmentation”, specifically for the load forecasting sub-module. Finally, IANOS iVPP Forecasting Engine provides predictions for load, generation and prices, for the Centralized Dispatcher described in T4.4 “Optimized cross-resource VPP coordination for energy services provision”.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 957810

3 Literature review and related work

While exploring new methodologies for energy forecasting it is vital to initially review the literature for related works. The focus of the current section is the presentation of the related implementations for load forecasting, PV and wind generation forecasting and finally energy market forecasting. Based on the work done in these implementations IANOS Forecasting Engine is going to further progress beyond the State-of-Art, using novel forecasting strategies or algorithms.

3.1 Time series forecasting

The data utilized in IANOS Forecasting Engine belong to a very specific category of datasets that are called time series, which have a unique way of being analyzed and processed. Time series are the set of data that express the evolutions of a variable during consecutive time periods, for example a day, a week, or a month and are also collected over a certain period of time. In time series, time is more often the independent variable and the goal is to make predictions for future times.

In order to predict the future values in energy related time series, different kinds of models are utilized, separated in basic categories, as presented in Figure 3.1. The first categorization is between analytical or physical models and data driven models. Analytical models are used to describe physical relationships between the values of the time series (e.g., station power generation) and relative parameters (e.g,. weather conditions, solar radiation). In the data driven models, the physical process is not taken into account. Data driven models are separated into: statistical models, machine learning models and hybrid models. The statistical models use mathematical models that derive from the correlation of the historical data with relative parameters, in order to find a function to best depict the behavior of time series. Machine learning models try to derive a relation between the relative parameters and values of the time series, while automatically improving its performance with experience. Lastly, hybrid models are combinations of statistical and machine learning models. In this document, the focus will be mainly on physical models and machine learning data driven models.

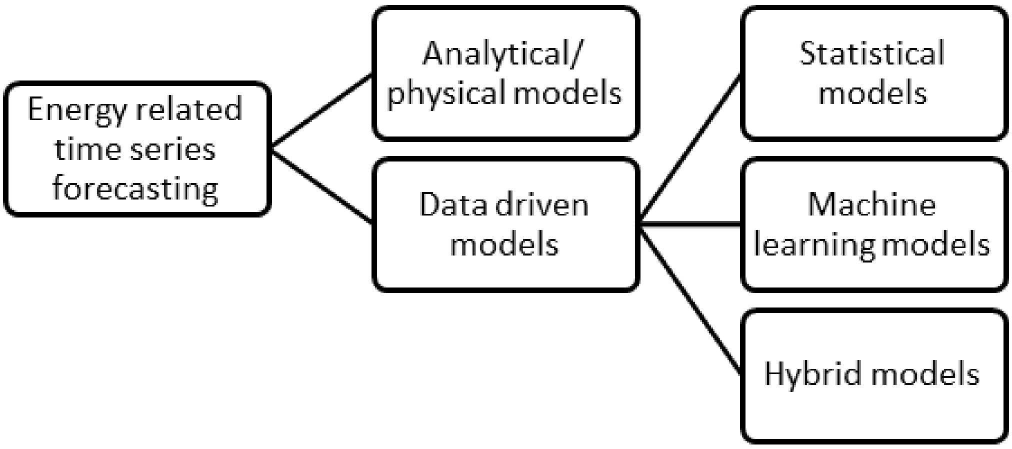

Figure 3.1: Classification of energy related time series forecasting

There are four types of time series forecasting, based on the prediction horizon, those are very short-term, short-term, mid-term and long-term forecasting. After the prediction, the results are evaluated, based on their comparison with the actual values, with the most common evaluation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

metrics being: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE or sMAPE).

3.2 Load forecasting

In smart electricity grids that depend heavily on renewable energy it is very important for the energy distributors to know in advance the energy requirements of the consumers. The overall process of predicting the amount of the energy required by consumers in order to fulfill their energy demands is described with the term load forecasting. Due to its importance, load forecasting has been the main focus of the research in recent years. In this chapter, works related to load forecasting, mainly using machine learning algorithms, are going to be presented. In addition, Table 3.1 lists the papers that are presented in this chapter.

In [1], Lahouar et al., Random Forests regression models are utilized in a real load data set from Tunisian Power Company for load prediction. They use features related to season, temperature (which was considered the most influential variable) and type of day to offer an accurate method with low errors for long test periods, without requiring an optimization process. Fumo and Biswas in [2] make use of simple and multiple linear regression models on a research house consumption dataset with daily and hourly data, showing improved results with features such as outdoor temperature and solar radiation. In [3] the goal is the prediction of short-term household's hourly energy consumption using Neural Gas, Decision Trees, MLP and XGBoosting models, while firstly using clustering methodologies to aggregate the consumers. Amin et al. in [4] make predictions with the use of ensembles of linear regression models in residential apartments dataset, while additionally clustering the usage profiles, demonstrating the significance of different time scales and a small percentage error. Short-term load forecasting is the objective of [5], where LightGBM, XGBoosting, and MLP are used on a power supply industry from Malaysia and wholesale load dataset from New England, with LightGBM giving the best results.

Tian et al. in [6] propose a framework based on the combination of CNN and LSTM networks. The framework is tested on electric load datasets provided by ENTSO-e Transparency Platform [7] for north Italy, and it is proved to outperform simple CNN and LSTM models. In [8] the researching team proposes a method using recurrence plots and two-dimensional CNN in a household dataset, where the time series dataset is encoded to recurrence plot images. The proposed model performs better than SVM, ANN and one-dimensional CNN. Almalaq and Edwards in [9] make a comparison between recurrent artificial neural networks, namely LSTM and GRU, and non-recursive networks, such as RBF Network and MLP for short term and medium test load forecasting in two aggregated energy consumption datasets for buildings in the US. They prove that the recursive models outperform non-recursive models. Ploysuwan in [10] utilize a combination of deep CNN as feature learning and LSTM network using data derived from IoT sensors from a household in Belgium and compare the model with GBM, Radial kernel SVM and Random Forest regressors, showing that the proposed model outperforms the previous comparison methods. Finally, Shabbir et al. in [11] use LSTM network in data collected from an Estonian household to achieve better results than shallow learning methods, such as Linear Regression, Tree-Based Regression and SVM.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Table 3.1: Load forecasting literature review

Work	Data	Algos used	Findings
[1]	Real load data set from Tunisian Power Company	Random Forest	Features related to season, temperature and type of day to offer an accurate method with low errors for long test periods, without requiring an optimization process
[2]	Research house consumption dataset with daily and hourly data	Multiple linear regression models	Improved results with features such as outdoor temperature and solar radiation
[3]	Short-term household's hourly energy consumption data	Neural Gas, Decision Trees, MLP and XGBoosting models	Better results when first use clustering methodologies to aggregate the consumers
[4]	Residential apartments dataset	Ensembles of linear regression models	The significance of different time scales and a small percentage error
[5]	Load dataset of a power supply industry from Malaysia and from New England	LightGBM, XGBoosting, and MLP	LightGBM gives the best results.
[6]	Electric load datasets provided by ENTSO-e Transparency Platform for north Italy	Framework based on the combination of CNN and LSTM networks	Proposed framework outperforms simple CNN and LSTM models
[8]	Household consumption dataset	Recurrence plots and two-dimensional CNN, where the time series dataset is encoded to recurrence plot images	The proposed model performed better than SVM, ANN and one-dimensional CNN
[9]	Two aggregated energy consumption datasets for buildings in the US	Recursive: LSTM and GRU Non-recursive: RBF and MLP	The recursive models outperformed non-recursive models
[10]	Data derived from IoT sensors from a household in Belgium	Combination of deep CNN as feature learning and LSTM network	The proposed model outperforms with GBM, Radial kernel SVM and Random Forest regressors
[11]	Data collected from an Estonian household	LSTM network	Model achieves better results than shallow learning methods, such as Linear Regression, Tree-Based Regression and SVM

3.3 Generation forecasting

In the IANOS project, the main source of energy generation comes from renewable energy sources. The use of renewable energy introduces uncertainty and fluctuations in the energy grid,.Therefore, since the start of the 21st century, a growth in research of renewable energy forecasting has been observed. In the current chapter, an analysis of literature related to renewable energy generation is presented, for both solar and wind power. Table 3.2 presents the reviewed works related to solar generation forecasting and Table 3.3 the works related to wind generation forecasting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Regarding the IANOS Forecasting Engine, in two pilot islands, Ameland and Terceira, the focus is on the prediction of energy generation from PV farms in both islands, and wind farms in Terceira.

3.3.1 Solar power (PV)

In the recent literature, both physical models and data driven models are used for solar power forecasting. Physical models are mainly used as a benchmark in forecasting tools, being already tested in various datasets. [12], [13]

The data driven models have been the main focus of research for solar power forecasting. These models consist of statistical and machine learning based models, with the latter being the main focus of the literature review implemented within the IANOS project. The literature provides an abundance of works that use machine-learning techniques for solar energy forecasting. Starting with [14], Zeng and Qiao, use SVM models in solar data collected from the National Solar Radiation Database, to predict the solar power. They prove that SVM models outperform statistical AR models. Random forest and multiple linear regression are used, by Huang et al., in [15] to predict the daily solar irradiance. The data used are Global horizontal solar irradiance data, was collected between 2003 and 2012. They found that the random forest model outperforms the multiple linear regression model, which has better results than the simple persistence models. Zhang et al in [16] utilize several shallow learning algorithms for day-ahead solar power forecasting, varying from RBFNN, SVM, kNN and weighted kNN. The dataset used in this work is derived from data from three different PV systems locations around the world, while using different features combinations, varying from solar irradiance, cloud coverance and more.

Lee et al., in [16], used a RNN model with LSTM to predict the power generation of photovoltaic stations, using data from meteorological stations, sensors and photovoltaic generation data. They found that those models are able to provide fast results with a high accuracy. In [18], Lee et al., collected data and used it to forecast solar power generation using LSTM and Convolutional Neural Networks. The data consist of date, time, power generation, irradiation, and temperature from 71 photovoltaic inverters and the findings were that the models used outperforms traditional regressors and deep neural networks. Finally, in [19] the research team make use of CNN, LSTM and their combination in 4 years of solar data, with the hybrid model outperforming the other models, but has the longest time.

Di Su et al. in [20] test several machine learning algorithms, namely Back Propagation Neural Network (BPNN), Elman Neural Network (ENN), Generalized Regression Neural Network (GRNN), Adaptive Network based Fuzzy Inference System (ANFIS), Nonlinear Autoregressive Neural Network with Exogenous Inputs (NARXNN), K Nearest Neighbours (kNN), Extreme Learning Machine (ELM) and Random Forest (RF), for solar power forecasting on a dataset of a PV plant in the UK and propose a model based on the weighted average of the best performing models. In [21], Leva et al. utilize a Hybrid Artificial Neural Network system based on MLP, hybridised with the physical Clear Sky Solar Radiation (SCRM) algorithm in data concerning PV power measured in the SolarTechLab in Politecnico di Milano, achieving good results, but with high time complexity.

Table 3.2: PV generation forecasting literature review

Work	Data	Algos used	Findings
[12] [13]	Various solar generation datasets	PVLib Physical models	Comparison of forecasts using physical models

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Work	Data	Algos used	Findings
[14]	Solar data from National Solar Radiation Database	SVM	SVM model outperforms AR
[15]	Global horizontal solar irradiance between 2003 and 2012	Random forest and Multiple linear regression	Random forest better than multiple linear regression better than simple persistence models
[16]	Data from three different PV systems locations around the world, while	RBFNN, SVM, kNN and weighted kNN	Better results using different features combinations, varying from solar irradiance, cloud coverage and more
[17]	Photovoltaic generation, sensor, and weather data	RNN and LSTM	Fast results, high accuracy
[18]	Date, time, power generation, irradiation, and temperature data from 71 photovoltaic inverters	LSTM and Convolutional Neural Networks	Model outperforms traditional regressors and deep neural networks
[19]	4 years of solar data	CNN, LSTM and their combination	The hybrid model outperforming the other models, but has the longest time.
[20]	Dataset of a PV plant in the UK	BPNN, ENN, GRNN, ANFIS, NARXNN, kNN, ELM and RF	Model based on the weighted average of the best performing models outperforms other models
[21]	PV power data measured in the SolarTechLab in Politecnico di Milano	Hybrid Artificial Neural Network system based on MLP, hybridised with the physical Clear Sky Solar Radiation (SCRM) algorithm	Good results, but with high time complexity

3.3.2 Wind power

The wind power forecasting problem has been proved to be more challenging than the solar power forecasting in the recent literature, due to the highest complexity of the weather features, namely wind speed and direction, which affect the power output of the wind farm. The models used vary from statistical, shallow learning, including Random Forest, Gradient Boosted, XGBoosting regressions or SVM, to more complex deep learning models, while ensembles of the previously mentioned methods are also utilized. [22]

In [23] Barbounis et al. make use of various RNN models to predict the power output of a wind park in Crete, Greece using meteorological predictions as input features. ANN and kNN models are used by Jursa and Rohrig in [24] to predict the power output of wind farms. The data used consist of weather data and historic power data of wind farms and the finding is that the models used achieve an improvement of around 11% over the persistent method. Short-term prediction of wind power is achieved by Jursa in [25] using several methods, such as ANN, SVM, kNN, and ensembles and power output data of German wind farms and weather prediction data from German weather service.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

The combinations of the models are found to produce better results. In [26] Click or tap here to enter text.Chen et al. use LSTM and SVRM, which are combined to an EnsemLSTM, to predict the wind speed using data from wind farms in China. They found that the ensemble method outperforms SVR, ANN, and kNN.

Table 3.3: Wind generation forecasting literature review

Work	Data	Algos used	Findings
[22]	Various wind power generation datasets	Statistical, Random Forest, Gradient Boosted, XGBoosting, SVM, and deep learning models	Comparisons between the models
[23]	Power output of a wind park in Crete, Greece	RNN models	Meteorological predictions as input features have better results
[24]	Weather data and historic power data of wind farms	ANN and kNN	Improvement of around 11% over the persistent method
[25]	Power output data of German wind farms and weather prediction data from German weather service	ANN, SVM, kNN, and ensembles	The combinations of the models are found to produce better results
[26]	Wind speed data from wind farms in China	LSTM and SVRM, combined to an EnsemLSTM	the ensemble method outperforms SVR, ANN, and kNN

3.4 Energy market forecasting

3.4.1 Energy markets in Europe

Since the start of the 21st century, the electricity system context in Europe went through significant changes, mainly due to the growing penetration of intermittent renewable energy sources and liberalization of energy markets. These factors have led to the need for a more efficient and lower-carbon system. EU Wholesale markets now experience the inclusion of a vast amount of distributed energy resources such as EVs and energy storage systems, while prosumers are having a key role into this energy transition. In IANOS, a need has appeared for accurate energy market forecasting together with generation and demand forecasting (focused mainly in the NL wholesale market including Ameland). More specifically, the markets that are further researched are the day-ahead market, the intra-day market, the imbalance market and the frequency response services market.

3.4.1.1 Day-ahead market

The day-ahead market concerns the bidding process that takes place on day D and with which the stakeholders make commitments for selling or buying specific amounts of electricity at each hour of day D+1. The price of energy is not already known at the time of the bidding but is formed when all the bids from the energy generators are revealed. The outcome of the day-ahead market clearing mechanism is the amount of the energy that each participant sells or buys and the price of the electricity. More specifically, each hour an amount and a price of energy is determined.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

3.4.1.2 Intra-day market

The bids that each participant submits in the day-ahead market have a high degree of uncertainty, mainly due to the uncertainties of renewable energy sources. To tackle this challenge more accurately, the mechanism of intra-day market has been introduced. According to this mechanism, the participants are able to make adjustments in the transaction of energy that are committed from the day-ahead process closer to the actual time of energy delivery, and thus while having more certain information for the availability of the assets and needs.

3.4.1.3 Balancing Market

Balancing energy market

The last main market of the modern electricity market system in Europe is the imbalance market or balancing energy market. This market is executed every hour of each day in order to handle the deviations of the commitments of the previously mentioned markets and the actual demand and generation in near real-time. With this mechanism, the price of the deviation of the power agent is determined.

Balancing capacity market

In order for the electricity system to function appropriately, the system operator must ensure that the frequency is set to a referenced value, namely at 50 Hz in Europe. To achieve that the operators use instruments, called frequency response services. The system outlined by the European Network of Transmission System Operators for Electricity (ENTSO-E) suggests that these services are divided into three main instruments, namely Frequency Containment Reserve (FCR), also called primary control reserve, automatic Frequency Restoration Reserve (aFRR), also called secondary control reserve and manual Frequency Restoration Reserve (mFRR), or tertiary control reserve.

The main difference between these mechanisms is the time of activation. More specifically, FCR reacts almost instantaneously, and must fix the deviation within 30 seconds in an automatic and nonselective manner. A certain amount of primary control is, at any given time, ready for activation, in order to ensure grid's stability. The secondary control reserve, aFRR, gradually replaces the FCR, if the deviation from the reference value of the frequency persists, and must provide reserve within 5 minutes. Finally, the mFRR must be fully deployable to assist or substitute the aFRR after 12.5 minutes.

3.4.2 Energy market forecasting review

In the past years, the focus of the research in energy market forecasting has mainly shifted to electricity price forecasting, for the three main markets. The importance of forecasting the electricity prices, namely day-ahead and intra-day, is highlighted in [27], where statistical methods are used to predict the electricity price in order to benefit a small-scale RES generator. Most of the works reviewed within the IANOS project make use of shallow or deep learning techniques, or combinations of the previous, to predict the future values of the prices. Table 3.4 lists the different papers that are presented.

In [28] the research team make use of SVM forecasting model in market clearing price, proving the model to be more robust and reliable than ANN and traditional approaches. An alternative to SVM, the informative vector machine (IVM) is used by Elattar in [29] to forecast the day-ahead price of electricity, while kernel principal component analysis is used to extract features, in real world datasets. González et al. in [30] highlight the effectiveness of an ensemble of regression tree

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

models, such as Bagging and Random Forests in electricity price dataset for the Spanish market. Electricity locational marginal price of New England electricity market is the forecasted variable of [31] where the researching team use an ensemble of XGBoost and relevance vector machines, with accurate and low time complexity results.

Recurrent Neural Networks, and more specifically LSTM and GRU are used in [32] to forecast the electricity prices time series of the Turkish day-ahead market, with GRU giving the best results. Different RNN models are also used from Anbazhagan and Kumarappan in [33] in prices dataset from the electricity market of Spain and New York. Finally, in [34] Kraft et al. use statistical models and ANN models to forecast the prices of the European FCR market, produce high forecast quality results.

Table 3.4: Energy market forecasting literature review.

Work	Data	Algos used	Findings
[27]	Day-Ahead and Intra-day price data	Statistical methods	Importance of forecasting electricity prices in order to benefit a small-scale RES generator
[28]	Market clearing price	SVM	Proposed model is more robust and reliable than ANN and traditional approaches
[29]	Real world datasets of day-ahead price of electricity	Informative vector machine (IVM)	Kernel principal component analysis is used to extract features
[30]	Electricity price dataset for the Spanish market	Ensemble of regression tree models, such as Bagging and Random Forests	Highlight the effectiveness of the proposed models
[31]	Electricity locational marginal price of New England electricity market	Ensemble of XGBoost and relevance vector machines	Accurate and low time complexity results
[32]	Electricity prices time series of the Turkish day-ahead market	LSTM and GRU	GRU gives the best results
[33]	Prices dataset from the electricity market of Spain and New York	RNN models	Highlight the effectiveness of the proposed models
[34]	Prices of the European FCR market	Statistical models and ANN models	High forecast quality results

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

4 IANOS Forecasting Engine

The Forecasting Engine is a key part of IANOS architecture. In this section, a view of the component's allocation is presented, together with an analysis of the different subcomponents and the connections to other submodules. Finally, the data model utilized within the Forecasting Engine is presented.

4.1 Conceptual View

IANOS Forecasting Engine is responsible for providing energy & price related forecasts to other components of IANOS system, while communicating with external components to retrieve necessary input data.

The Forecasting Engine is connected with IANOS Secure Enterprise Service Bus (ESB) to receive generation and consumption data and numerical weather predictions, while external APIs provide energy market data. The engine will utilize the clustering results of Aggregation and Classification module in order to facilitate load forecast based on the cluster of each prosumer. It will receive as input the label of the cluster that the consumption profile of each consumer belongs. Currently, the models that have been created are single-prosumer based, but the aim is to support both personalized and cluster based models.

The forecasts are retrieved by the Centralized Dispatcher to optimally dispatching field-level IANOS elements and the DLT-based Transactive Logic in order to facilitate direct energy transactions in the community. The following figure presents the allocation of the Forecasting Engine into IANOS architecture component and its interaction with other modules.

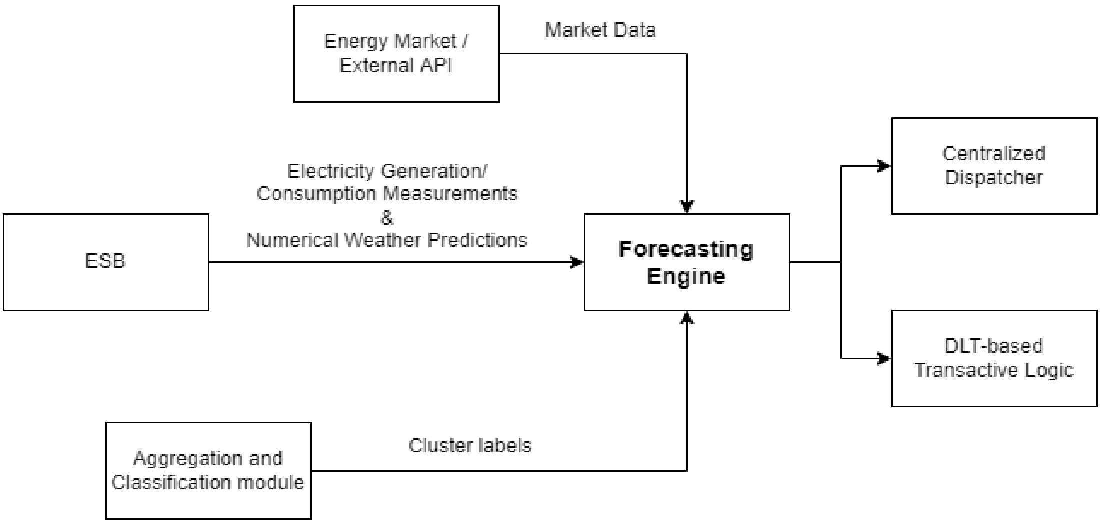

Figure 4.1: Allocation of the forecasting component into the IANOS project's architecture

4.2 Components and outer dependencies

IANOS Forecasting Engine consists of four different sub-components, namely Demand Forecasting, PV Generation Forecasting, Wind Generation Forecasting, and finally Energy Market Forecasting. For each of the subcomponents, there is a dependency on data retrieved from other components or external sources. More specifically, historical data, regarding energy generation and consumption, and numerical weather predictions will be stored in a time series database and will be accessible

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

through ESB, cluster labels need to be provided by Aggregation and Classification module and finally price data and generation and demand forecasts are provided by ENSTO-E [7] energy market. Depending on what is requested, the final output of the component is either the day-ahead, or the intra-day forecast. A detailed functionality diagram of IANOS Forecasting Engine can be observed in the figure below.

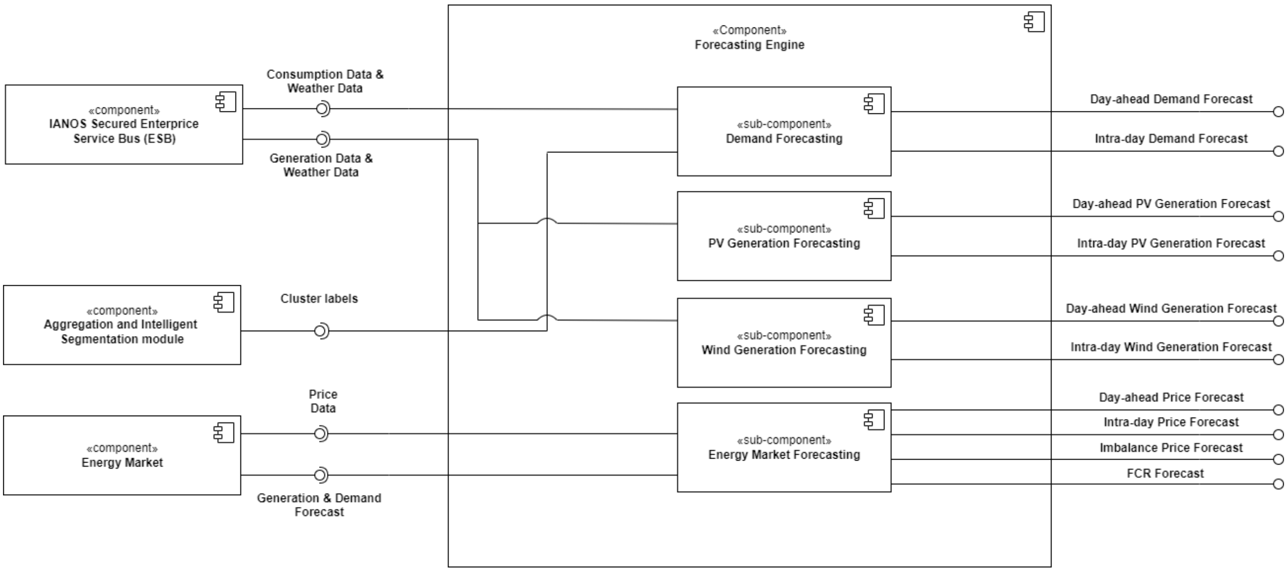

Figure 4.2: Detailed functionality diagram of IANOS Forecasting Engine

4.3 Data model

The outputs of IANOS Forecasting Engine follow a specified data model dependent on the resolution of the forecast. For both categories of the forecast, the engine outputs a .json file containing information for the id of the prosumer and the asset, for which the predictions are made, and information about the forecast, namely the type of forecast, the horizon of forecast and the measurement unit. Additionally, for intra-day forecasts under the properties key the timestamp and the value of the corresponding prediction is depicted. Finally, for day-head forecast there are either 24 timestamp-value pairs for hourly resolution datasets, or 96 timestamp-value pairs for quarterly resolution datasets.

In the following figures, there are examples of the outputs of the Forecasting Engine for intra-day and day-ahead forecasts. Additionally, Table 4.1 is provided to illustrate descriptions of the inputs and output variables of the Forecasting Engine

Table 4.1: Description of the input and output variables of the Forecasting Engine

Friendly name	**Measurement name**	**Measurement unit**	**Measurement type**	**Description**
Residential Load Forecasting Input				
Power Consumption	active_power	W	float	Measurement name of the respective historical load measurements
Temperature	temp	Celsius	float	Historical values of temperature predictions

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Friendly name	Measurement name	Measurement unit	Measurement type	Description
Residential Load Forecasting Output				
Forecasted Energy Consumption	energy	Wh	float	The forecasted value of load
Solar Generation Forecasting Input				
Power Generation	active_power	W	float	Historical values of solar generation
Temperature	temp	Celsius	float	Historical values of temperature predictions
Cloud Coverage	cloud_cover	-	float	Historical values of cloud coverage predictions
Wind Speed	wind_spd	m/s	float	Historical values of the wind speed predictions
Solar Generation Forecasting Output				
Forecasted Solar Generation	energy	Wh	float	The forecasted value of solar generation
Wind Generation Forecasting Input				
Power Generation	active_power	W	float	Historical values of wind generation
Temperature	temp	Celsius	float	Historical values of temperature predictions
Cloud Coverage	cloud_cover	-	float	Historical values of cloud coverage predictions
Wind Speed	wind_spd	m/s	float	Historical values of the wind speed predictions
Wind Direction	wind_dr	degrees from north	float	Historical values of wind direction predictions
Wind Generation Forecasting Output				
Forecasted Wind Generation	energy	Wh	float	The forecasted value of wind generation
Energy Market Forecasting Input				
Price or FCR	da_price intra_price imb_price fcr	EUR/MWh EUR/MW/ISP	float	Historical values of price or FCR
Day ahead load forecast	da_load	MWh	float	Historical values of day ahead load predictions for all the region

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Friendly name	Measurement name	Measurement unit	Measurement type	Description
Day ahead generation forecast	da_gen	MWh	float	Historical values of day ahead generation predictions for all the region
Energy Market Forecasting Output				
Forecasted prices or FCR	da_price_for intra_price_for imb_price_for fcr_for	EUR/MWh EUR/MW/ISP	float	Forecasted values of price or FCR


```
{
  "asset_id": "bdksncjdsijcd",
  "measurements": [
    {
      "measurement_name": "activePower",
      "measurement_unit": "kW",
      "measurement_type": "float",
      "properties": [
        {
          "value": 50.53363,
          "timestamp": "2019-11-08T12:45:40.035Z"
        },
        {
          "value": 60.53363,
          "timestamp": "2019-11-08T12:50:40.035Z"
        }
      ]
    },
    {
      "measurement_name": "reactivePower",
      "measurement_unit": "kVar",
      "measurement_type": "float",
      "properties": [
        {
          "value": 15.53363,
          "timestamp": "2019-11-08T12:45:40.035Z"
        },
        {
          "value": 35.53363,
          "timestamp": "2019-11-08T12:50:40.035Z"
        }
      ]
    }
  ]
}
```


Figure 4.3: Example of the input of the Forecasting Engine for day-ahead forecast.

4.4 Software pre-requirements

The reference language of the Forecasting Engine component, in order to facilitate its functionalities, will be **Python**[35]. Python is ideal for fast and easy prototyping and while containing suitable

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

packages for machine learning based tasks. More specifically, in the implementation of IANOS Forecasting Engine, the packages are used for data manipulation are **Numpy**[36], which offers an immense set of mathematical functions that enable fast and convenient array operations, and **Pandas**[37]**.** Pandas is a well-known and widely-used data analysis library. It is a simple tool for working with datasets in a variety of formats, including csv, h5, json and others. Pandas is great for time-series analysis since it includes a variety of ways for reading, writing, and pre-processing raw time-series data by extracting key data.

In the development of IANOS Forecasting Engine, for data modeling, there are two main packages used, namely **Scikit-learn**[38], that offers various pre-built supervised machine learning algorithms, including the tree-based models and SVM that are utilized within IANOS Forecasting Engine, and **Keras**[39] which provides Neural Network algorithms.

Lastly, the library utilized for visualizing the data and the results is **Plotly**[40], which provides useful tools for interactive graphics and plots.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

5 IANOS forecasting methodology

In this section, the progress of IANOS Forecasting Engine beyond the state-of-art is described. The techniques used are focused on delivering lightweight forecasting agents that accurately forecast the consumption, generation and price of electrical energy in the two pilot islands.

The methodology used within the module is mainly separated into four steps as explained in Figure 5.1. The first step includes the acquisition of the data, in which data is collected either from local or external databases or by utilizing specific APIs that allow frameworks to communicate with each other. The next step is the preprocessing and cleaning of the data, whereas the module detects missing values and outliers and cleans or drops them respectively and finally normalizes the data. The first two steps are common to all the forecasting sub-modules and they are further described in the following chapters.

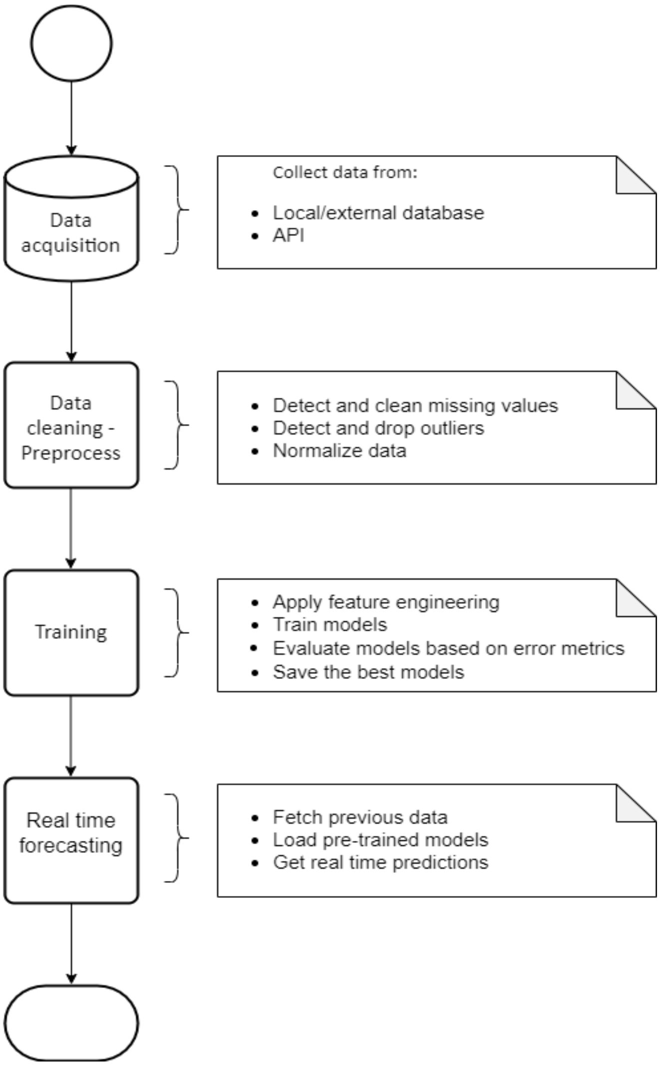

Figure 5.1: IANOS methodology flowchart

The third step of the methodology is the training of the models used in the Forecasting Engine, in which the feature engineering is carried out as well. Based on defined evaluation metrics, the models with the best performance are identified and saved for future use. For each submodule, the features and the algorithms used may vary, but the logic remains. The final step of methodology is real time

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

forecasting, where the aforementioned pre-trained models are loaded and the previous data is fetched in order to get real time predictions.

In energy related time series it is possible to encounter changes in the incoming measurements that may lead to poor forecasting performance. To facilitate this kind of problem the models are frequently re-trained in specific time periods, aiming to adapt in the most recent conditions.

A separate note has to be made for load forecasting where the forecast methodology is going to be held with respect to the cluster that each consumer belongs. The Forecasting Engine will acquire the clustering labels for the consumption of each actor, and a model will be built based on the datasets of all the actors belonging to the same cluster. By following this methodology, it is easier to provide forecasts for newly installed consumers/prosumers simply by assigning them in the appropriate cluster. The Aggregation and Classification module developed within IANOS project is further described in Deliverable D4.5 “iVPP Aggregation and intelligent Segmentation”.

Based on the analysis of IANOS project’s different scenarios, there have been two main time horizons of forecasting, as presented in Figure 5.2: Forecast horizons considered in the energy forecasting model. In the day-ahead forecast, in the start of each day, the values of the whole next day are forecasted, while the short-term forecast will be executed on a specific time-intervals in order to generate the next value. In the different submodules the forecasting horizons may vary, thus a more detailed explanation is going to be presented in the following chapters.

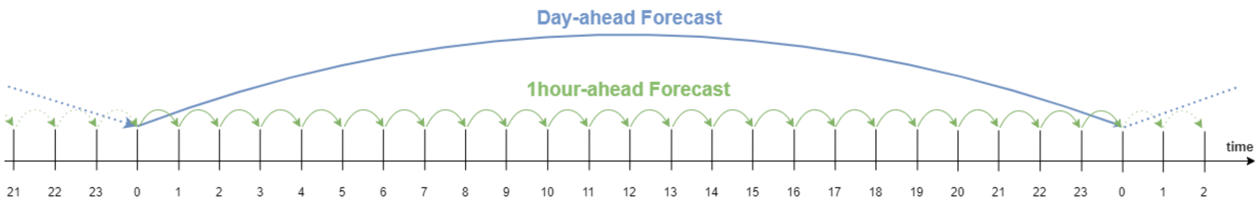

Figure 5.2: Forecast horizons considered in the energy forecasting model

5.1 Data collection

The IANOS Forecasting Engine tool requires proper data exchange between different system components, external components, and field devices. The data transfer/collection role is played by the iVPP secured Enterprise Service Bus (ESB), focusing specifically on cyber-security aspects. Within IANOS, the ESB will be the tool that all the components integrated within the iVPP platform will utilize to collect data of all the energy assets that are distributed in different geographical locations by communicating with it. ESB will provide specified methods to exchange the contextual data from the field components to the iVPP platform. For the collection of the data, secure password protected connection to the database is required. Finally, in order to retrieve external data, namely price data and weather predictions, on-line connections with external databases will be established through Application Programming Interfaces (API).

Finally, it must be mentioned that Task 4.1 “Cyber-Secure data monitoring and VPP Governance” will be responsible for highlight all the cybersecurity issues regarding the transactions through the ESB.

5.2 Data preprocessing and outliers detection methods

After the acquisition of the raw data from a database or an API, the next step of the IANOS Forecasting Engine’s methodology is the preprocess of the data. The raw data can be inaccurate or

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 957810

corrupted, with several missing values or outliers. This introduces unacceptable errors in the forecasting methodology and may result in inaccurate forecasting. Therefore, the module firstly checks the data for missing values and then replaces them using interpolation strategies that best fit to the form of the time series. Additionally, the module utilizes the z-score [41] function in order to remove the outlier points of the time series, to eliminate the bias. Finally, when needed, the dataset is scaled and normalized using the most common techniques, namely MinMaxScaler [42] or StandardScaler [43].

5.3 Load forecasting

5.3.1 Data exploration

In the current state of the pilot islands, there are no datasets available for load forecasting. Thus, the dataset used for the first version of load forecasting is a private electricity consumption dataset from a residential user in Thessaloniki, Greece. The measurement unit of dataset is Watt (W) and the resolution of measurements is quarterly. Additionally, numerical weather prediction that correspond to the time and place of consumption measurements are utilized as features. In the following figure Figure 5.3, there is an example plot of the energy consumption data of the residential user.

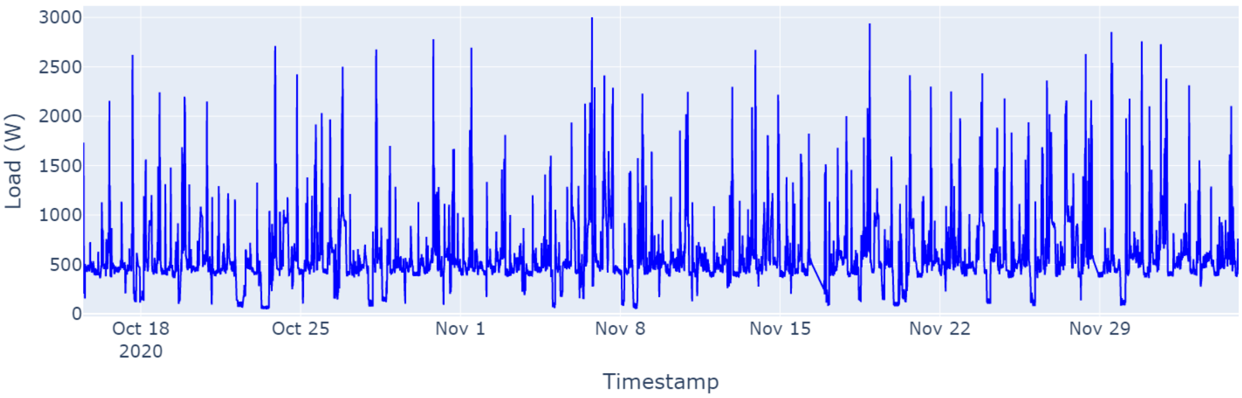

Figure 5.3: Energy consumption example plot for residential user.

The following table Table 5.1 provides necessary statistics for the final dataset used for the electricity load of the residential user.

Table 5.1: Load dataset's statistics

	Load dataset
Count	19257
Mean	14.19
Std. Deviation	7.24
Min	-4.4
Max	35.6
Start Date	2020-08-30 00:00:00

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

End Date	2021-03-18 12:30:00

5.3.2 Load forecasting strategy and feature creation

The module is able to provide multi-step forecasts with configurable horizons, in order to make day-ahead and intra-day forecasts. For these multi-step forecasts, separate models that are created to predict each time step of the horizon. The horizon of the forecast depends on the requested target, which within the IANOS project corresponds to intra-day and day-ahead forecasts. This direct strategy for multi-step forecasts and the corresponding features used are highlighted in Figure 5.4: Load forecasting strategy.

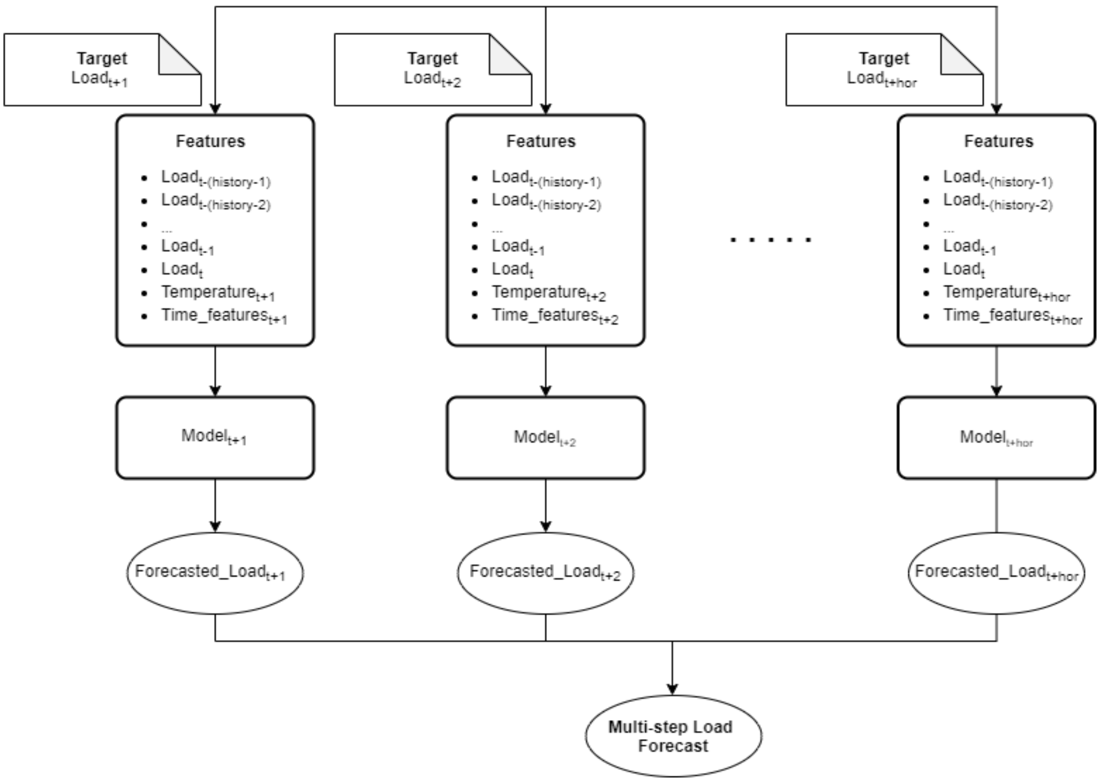

Figure 5.4: Load forecasting strategy

The features utilized as input to the data-driven models can be separated into three main categories, namely energy based features, which are the historical consumption values gathered from the energy smart meters, weather features, which are retrieved from an external API, and temporal features to capture the season periodicity.

More specifically, the energy based features consist of past load values within a previously specified time window, named history. Thus, in order to predict the t+1 load, the models utilize previous load values from t timestep to t-(history-1) timestep. In IANOS forecasting module, the history is set to 24 hours or 96 quarters depending on the resolution of the time series.

The weather features used for the load consist of only the temperature feature, as this is the only feature proven useful in experimentations. For the multi-step forecast, the temperature of each time-step is used in the corresponding model.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Lastly, the time features are used with the same logic as the weather features. In order to capture the time seasonality of the load time series, temporal cyclical features are used through the following equations:

$$sin_{min} = sinus\left(2\pi * \frac{minutes}{60}\right) \quad (1) \qquad cos_{min} = cosinus\left(2\pi * \frac{minutes}{60}\right) \quad (2)$$

$$sin_{hour} = sinus\left(2\pi * \frac{hour}{24}\right) \quad (3) \qquad cos_{hour} = cosinus\left(2\pi * \frac{hour}{24}\right) \quad (4)$$

$$sin_{wday} = sinus\left(2\pi * \frac{day\_of\_week}{6}\right) \quad (5) \qquad cos_{wday} = cosinus\left(2\pi * \frac{day\_of\_week}{6}\right) \quad (6)$$

$$sin_{yweek} = sinus\left(2\pi * \frac{week\_of\_year}{54}\right) \quad (7) \qquad cos_{yweek} = cosinus\left(2\pi * \frac{week\_of\_year}{54}\right) \quad (8)$$

In addition to the temporal cyclical feature, categorical features indicating the part of the day (e.g. morning, noon, etc), Function 9, and part of the week (e.g. workday, weekend), Function 10, were extracted from each timestamp of the time series.

$$day\_part = \begin{cases} dawn, & if(hour \geq 2)\ AND\ (hour \leq 5) \\ morning, & if(hour \geq 6)\ AND\ (hour \leq 9) \\ noon, & if(hour \geq 10)\ AND\ (hour \leq 13) \\ afternoon, & if(hour \geq 14)\ AND\ (hour \leq 17) \\ evening, & if(hour \geq 18)\ AND\ (hour \leq 21) \\ midnight, & if(hour \geq 22)\ OR\ (hour \leq 1) \end{cases} \quad (9)$$

$$is\_weekend = \begin{cases} 1, & where\ ((dayofweek = 5)\ OR\ (dayofweek = 6) \\ 0, & where\ ((dayofweek \neq 5)\ AND\ (dayofweek \neq 6) \end{cases} \quad (10)$$

5.3.3 Prediction Models

Many data-driven models were performed in order to find the most efficient in both forecast accuracy and time complexity. More specifically, the models utilized can be classified into the following categories, as shown in the Table 5.2:

Table 5.2: IANOS Forecasting Engine's prediction models

Category	Prediction model
Bagging trees ensemble	• Random Forests (RF)
Boosting trees ensembles	• Gradient Boosting Regressor (GBR) • eXtreme Gradient Boosting (XGBoost)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

	• Light Gradient Boosting Machine (LightGBM)
Support Vector Machines (SVM)	• Support Vector Regressor (SVR)
Artificial Neural Nerworks (ANN)	• Multi-Layer Perceptron (MLP) • Long-short Term Memory Recurrent NN (LSTM RNN)

The aforementioned models are also employed in the generation and energy market forecast modules, with parameters set according to the problem. In generation forecasting, physical models for both PV and wind generation are going to be implemented, in the following version of the deliverable.

5.4 PV forecasting

5.4.1 PV analytical/physical model

The physical model used in the IANOS Forecasting Engine for solar power forecasting is based on the open-source software that models the behavior of PV systems, PVlib[44]. PVlib is a procedural Python library used to simulate the performance of PV systems, using a set of functions and classes. The tool helps users model the outputs of a PV system by using different models and customizable PV system parameters that include location and time, PV system configuration characteristics and forecasted weather data, relevant to the solar power modeling.

As mentioned previously, the physical model is going to be developed in the second version of the deliverable due to the current lack of data.

5.4.2 PV data driven approach

In the data-driven approach for PV generation forecast, there are no requirements for the physical characteristics of the photovoltaic system, and only past values of the PV generation are used, together with time and weather features.

5.4.2.1 Data exploration

For the PV generation forecasting with data-driven models one dataset was used for the Ameland pilot island. In this first version of the deliverable there was a lack of data for the Terceira pilot. Additionally, weather forecast data were collected from external APIs that correspond to the respective time periods, for the pilot island.

The PV generation dataset used in generation forecasting for Ameland pilot island correspond to the solar power production of the installed 6MWp community owned solar farm. The overall production of the solar farm can be visualized in [45], while the data were retrieved from the source API url using a python script. The measurements come in 5-minute resolution, and were resampled in quarters, from 2020-11-30 to 2021-07-21, which corresponds to 22364 entries, with a few missing values that were interpolated. The measurement unit of the data is kW. An example of the dataset is shown on the following Figure 5.5.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

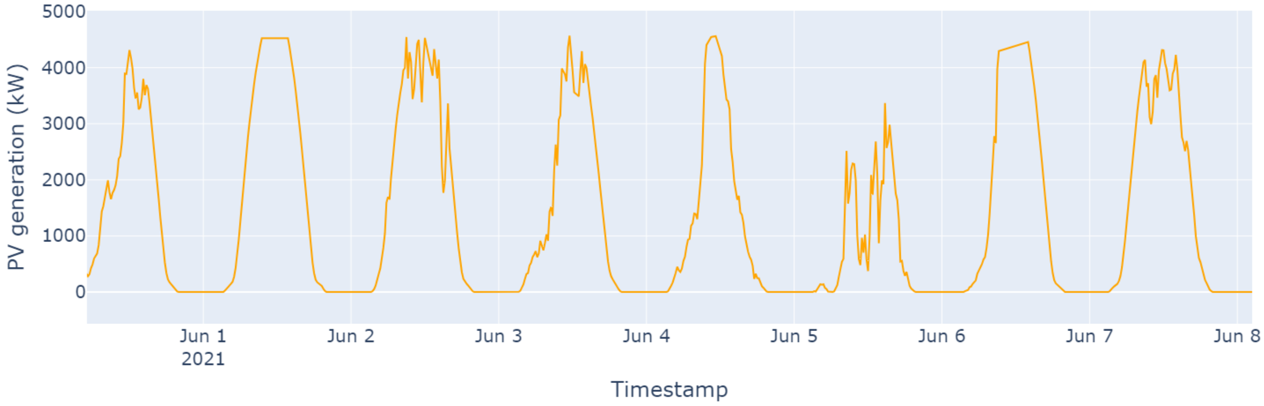

Figure 5.5: PV generation example plot for one week [Ameland]

The following table provides necessary statistics for the final dataset used for the PV generation forecasting for Ameland.

Table 5.3: PV generation dataset's statistics [Ameland]

	PV generation dataset
Count	22364
Mean	740.185
Std. Deviation	1276.384
Min	0.0
Max	4806.436
Start Date	2020-11-30
End Date	2021-07-21

5.4.2.2 PV forecasting strategy and feature creation

With the same logic as the load forecasting sub-module, the direct strategy of the multi-step PV forecast can be viewed in the Figure 5.6. The different forecast horizons correspond to intra-day and day-ahead PV generation forecasting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

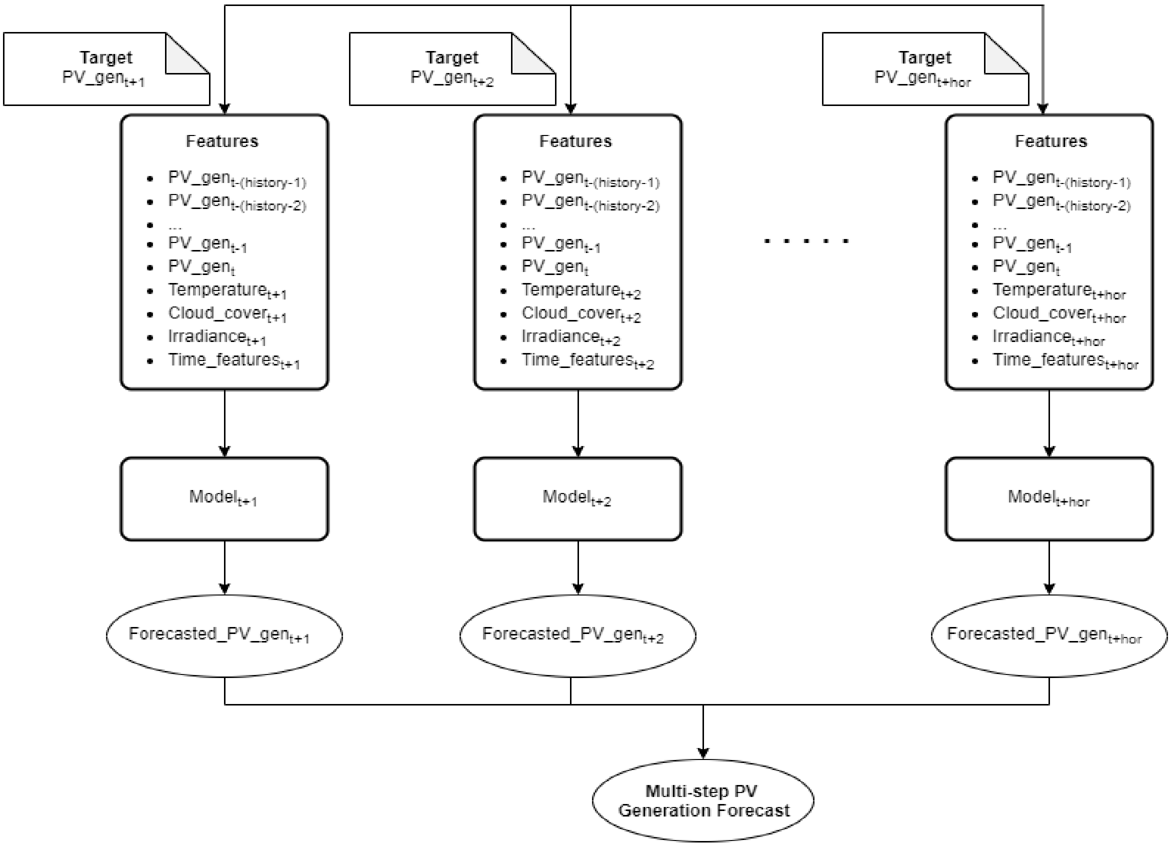

Figure 5.6: PV generation forecasting strategy

The main difference between PV and load forecast is the features used in each occasion. In PV generation forecasts, the past values of the PV farm generation are utilized as energy based features in a predefined time window. As far as weather features go, the ambient temperature, the cloud coverage and the solar irradiance are used. The time features used are the same as in load forecasting.

5.5 Wind forecasting

5.5.1 Wind analytical method

As mentioned previously, in the IANOS Forecasting Engine, both physical and data driven models are used for generation forecasting. More specifically, the physical model used for wind power forecasting is the open-source library Windpowerlib [46].

In the analytical method approach, the entire system is modeled with the use of equations that derive from wind turbine mechanics. The contributing factors that the wind power output is depended on can be grouped into two main categories:

- Weather variables: temperature, pressure, wind speed and wind direction
- Technical characteristics of the turbine: coefficient curves provided by the manufacturer, nominal power, rotor diameter and hub height.

These factors are required by the Windpowerlib for the calculation of the wind turbine's power output. The library defines the wind power [W] output using Function 11, in which v_{wind} corresponds to the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

wind speed [m/s] at the hub height, d_{rotor} is the rotor diameter [m] of the wind power, ρ_{hub} stands for the air density [kg/m^3] at hub height and $cp(v_{wind})$ are the coefficient curves:

$$P = \frac{1}{8} \cdot \rho_{hub} \cdot d_{rotor}^2 \cdot \pi \cdot v_{wind}^3 \cdot cp(v_{wind}) \quad (11)$$

In the same manner as for the PV forecasting, the physical model for wind forecasting is going to be developed in the second version of the deliverable due to the current lack of data.

5.5.2 Wind data driven approach

In the same manner as the PV data driven approach, in wind data driven approach there is no need for the physical characteristics of the wind farms, only the past generation values and additional features, namely weather and time features.

5.5.2.1 Data exploration

The wind power generation forecasting sub-component will be deployed in Terceira Pilot Island. Currently there are no available data for wind power modeling, thus a public dataset is used for the version of the deliverable. The dataset can be found in Kaggle [47]. It contains wind power measurements from a wind park together with weather forecasts, namely wind velocity in 10m and 100m above the ground. The wind power output is normalized by the maximum output of the wind farm. In the dataset there are close to two years of hourly wind power data. In Figure 5.7, there is an example plot of the dataset.

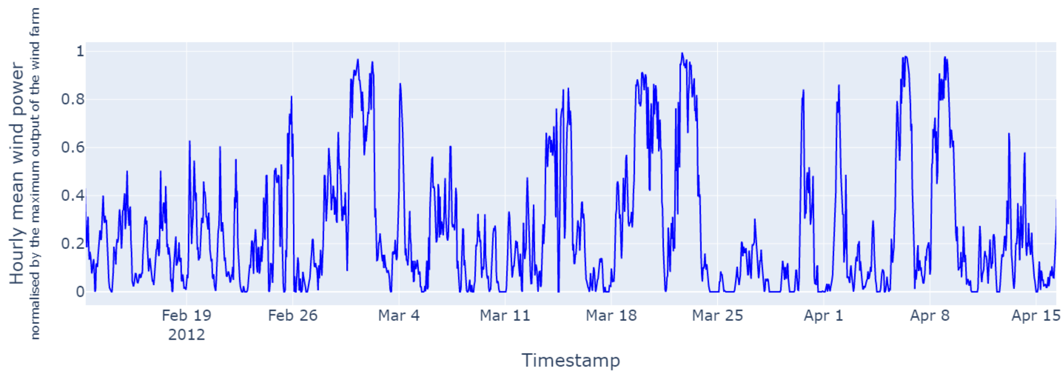

Figure 5.7: Wind power generation example plot

Additionally, in Table 5.4, the necessary statistics of the dataset are provided.

Table 5.4: Wind generation dataset's statistics

	Wind generation dataset
Count	16765
Mean	0.303
Std. Deviation	0.290
Min	0
Max	1.0
Start Date	2012-01-01 01:00:00
End Date	2013-11-30 00:00:00

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

5.5.2.2 Wind forecasting strategy and feature creation

In the data driven approach of wind generation the different forecast horizons correspond to intra-day and day-ahead wind generation forecasting. The strategy of the forecast is similar to load and PV generation forecast, as shown in the Figure 5.6

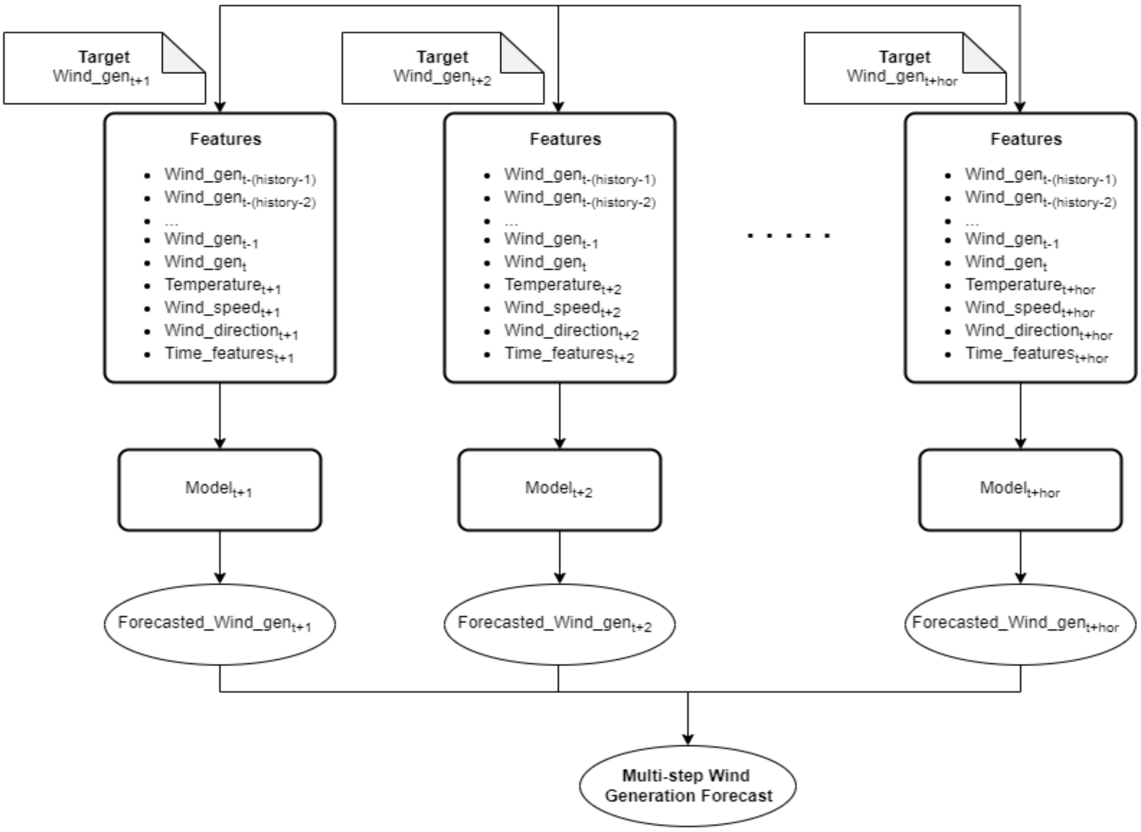

Figure 5.8: Wind generation forecasting strategy

The energy based features used in wind generation forecasts are the previous values of the wind generation of the wind farm in a previously specified time window. The weather features are the ambient temperature and the speed and direction of the wind. The time features used are the same as in load and PV generation forecasting.

5.6 Energy Market forecasting

The energy market forecasting tool developed within the IANOS project provides forecasts for the main components of the energy market. More specifically, forecasts for the day-ahead price, the intra-day price, the imbalance price and finally for the frequency containment reserve (FCR) are provided by the energy forecasting sub-module. The energy market forecasting tool will be applied in electricity prices and FCR concerning the energy market of the Netherlands.

5.6.1 Data exploration

The day-ahead, imbalance and FCR data utilized in the training of the energy market forecasting sub-module are collected via the ENTSO-E online platform [7]. European Network of Transmission System Operators (ENTSO-E). It represents 42 electricity TSOs from 35 countries across Europe.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

The platform provides various energy related data, varying from country load or generation profiles to energy market prices, utilized in this sub-module. The datasets provided by the ENTSO-E are usually relatively clean and have great quality overall.

Currently there is no available dataset for intra-day price forecasting, as ENTSO-E platform does not provide them. For this deliverable intra-day price dataset for the UK market is used, only for the initial experimentation due to lack of intra-day price data from Netherlands, which can be retrieved from [48].

In addition to the day-ahead electricity prices data, ENTSO-E provides day-ahead generation forecast data and day-ahead load forecast data for the whole country of the Netherlands. These datasets were also retrieved in order to be used as features for the energy market forecast, as they play a significant role in the fluctuations of the prices. These features are used in all of the sub-components of energy market forecasting.

5.6.1.1 Day-ahead price

The day-ahead price data for the Netherlands were collected during a one-year period, from 2020-08-31 to 2021-09-05, and the data are reported in EUR/kWh. The measurements come in hourly resolution, which make the total count of entries 8880, without any missing values. In the following figure, there is an example of Day-ahead price dataset.

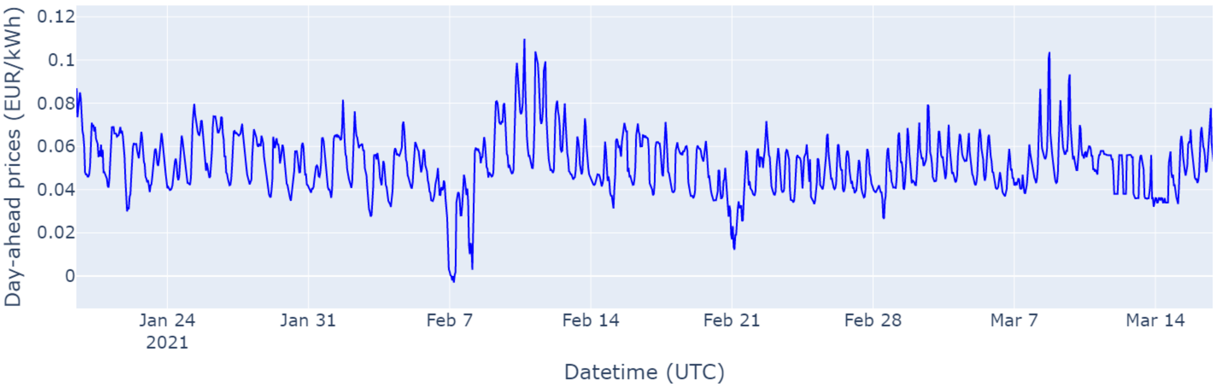

Figure 5.9: Day-ahead price example plot.

Table 5.5 provides necessary statistics for the final dataset used for the day-ahead price forecasting.

Table 5.5: Day-ahead price dataset's statistics.

	Day-ahead price dataset
Count	8880
Mean	0.058
Std. Deviation	0.024
Min	-0.066
Max	0.20
Start Date	2020-08-31 22:00:00

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

End Date	2021-09-05 21:00:00

5.6.1.2 Intra-day price

As mentioned previously the intra-day price dataset used in this deliverable correspond to UK intra-day prices. The measurement unit is EUR/kWh and the data have half-hourly resolution. In the following figure, there is an example of the intra-day price dataset, while the dataset's statistics are presented in Table 5.6.

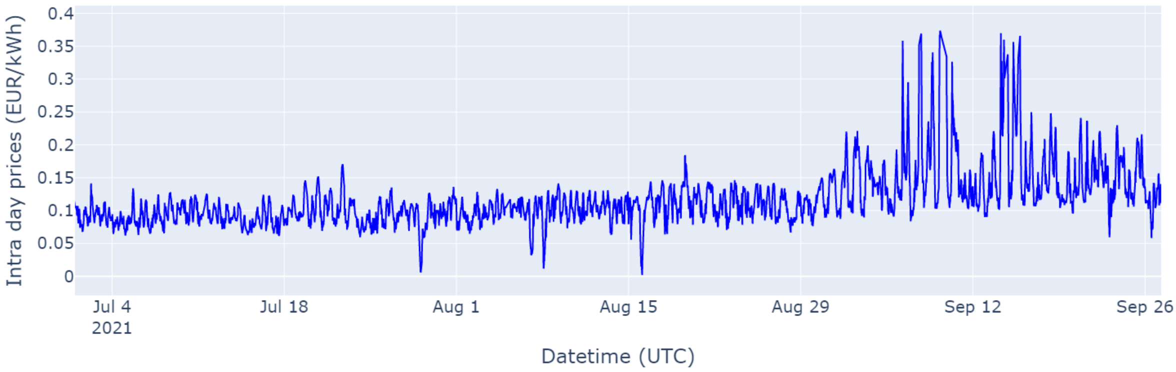

Figure 5.10: UK intra-day price example plot.

Table 5.6: UK intra-day price dataset's statistics.

	Intra-day price dataset
Count	7920
Mean	0.063
Std. Deviation	0.027
Min	-0.043
Max	0.373
Start Date	2021-01-01 00:00:00
End Date	2021-06-14 23:30:00

5.6.1.3 Imbalance price

In the same manner as the day-ahead prices, the imbalance price data are collected in the span of one year, from 2020-08-31 to 2021-09-05. The measurement unit of the data is EUR/MWh. The measurements come in quarterly resolution, which make the total count of entries 35516, without any missing values. In Figure 5.11 there is an example of imbalance prices.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

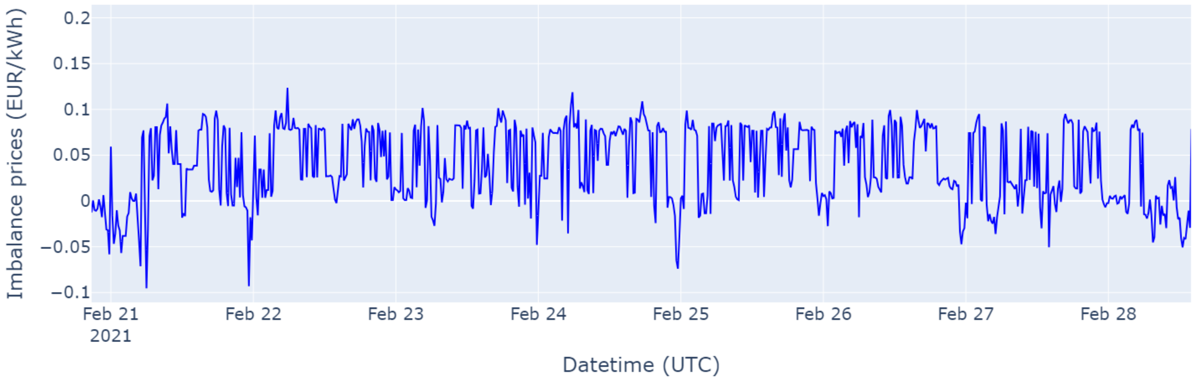

Figure 5.11: Imbalance price example plot.

In Table 5.7, necessary statistics for the final dataset used for the imbalance price forecasting are presented.

Table 5.7: Imbalance price dataset's statistics.

	Imbalance price dataset
Count	35516
Mean	0.056
Std. Deviation	0.072
Min	-2.198
Max	3.964
Start Date	2020-08-31 22:00:00
End Date	2021-09-05 20:45:00

5.6.1.4 FCR

The FCR dataset for Netherlands was collected between 2021-06-30 and 2021-10-01. The data available in ENTSO-E has resampled from daily prices to 4-hourly prices in May 2020 and the prices in the Netherlands were unusually high from mid 2020 to early 2021. However, in the three month time period that was elected, the prices for the FCR in the Netherlands have stabilized to the European level, which is the reason why this period was elected. The regulation price of FCR is estimated in EUR/MW/ISP, where ISP stands for imbalance settlement period. The measurements from the ENTSO-E come in hourly resolution, but they are the same for four hours, which makes the total count of entries 2232, without any missing values. In the following figure there is an example of the FCR dataset for August 2021.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

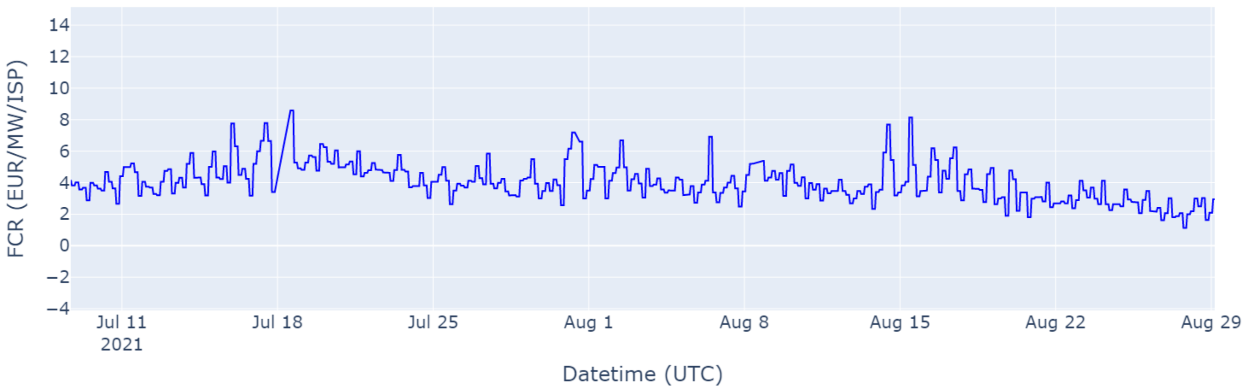

Figure 5.12: FCR example plot for one month.

Important statistics for the final dataset used for the FCR forecasting, are presented in the following table.

Table 5.8: FCR dataset's statistics.

	FCR dataset
Count	2232
Mean	4.186
Std. Deviation	1.625
Min	1.13
Max	14.27
Start Date	2021-06-30 22:00:00
End Date	2021-10-01 21:00:00

5.6.2 Energy market forecasting strategy and feature creation

In the same manner as in the aforementioned forecasting sub-modules the direct strategy of the energy market forecasting is adopted and illustrated on the following figure. The horizons of the forecasts are 24 hours for day-ahead price, 1 hour for intra-day price and FCR forecasting, and 1 quarter for imbalance price forecasting. The additional features used in energy market forecasts are, as mentioned previously, day-ahead generation forecast data and day-ahead load forecast data for the whole country of the Netherlands.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

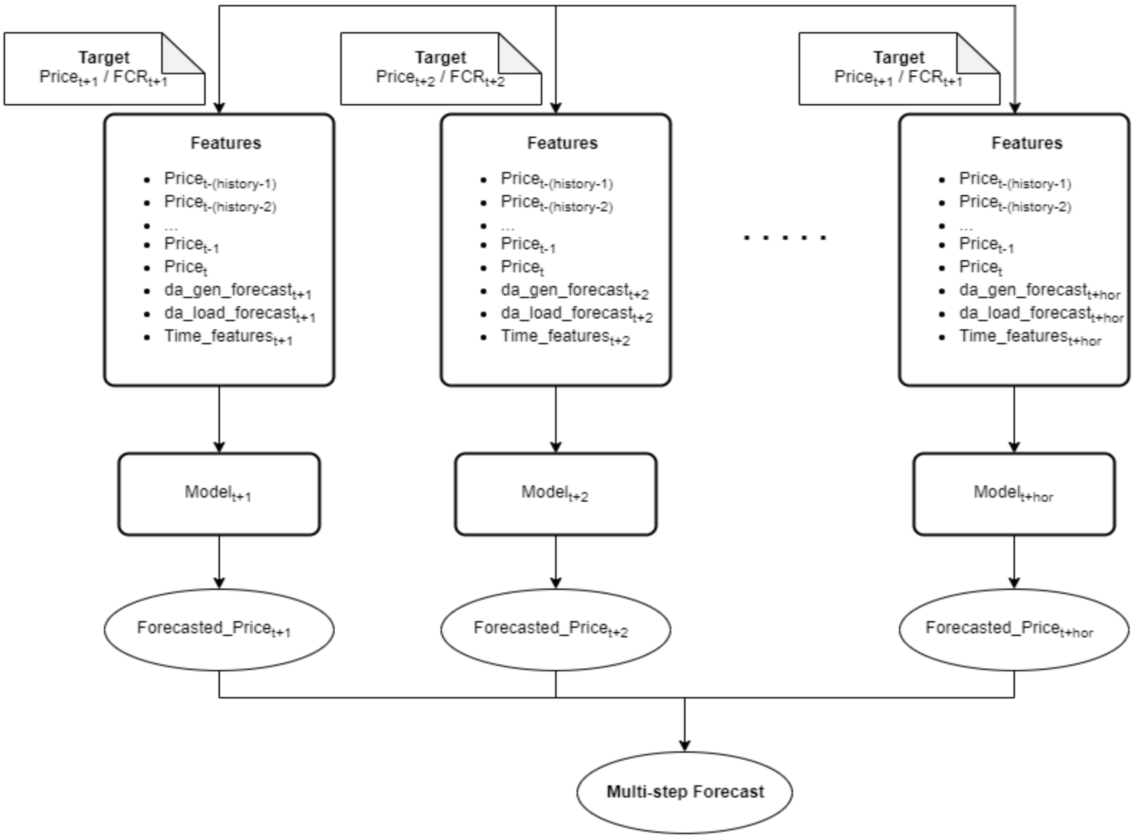

Figure 5.13: Wind market forecasting strategy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

6 Forecasting evaluation and results

6.1 Evaluation metrics

There were five different metrics used for the evaluation of the implemented methods. Table 6.1 presents the metrics and their respective formula, where, in each formula, A represents the actual value and F represents the forecasted value, their difference represents the error, and n represents the number of observations. Together with the following evaluation metrics, the execution time of the training procedure was estimated in each experiment.

Table 6.1: Evaluation metrics' formulas

Evaluation Metrics	Formulas
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^{n} \lvert A_i - F_i \rvert}{n} \quad (12)$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (A_i - F_i)^2}{n}} \quad (13)$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left\lvert \frac{A_i - F_i}{A_i} \right\rvert * 100\% \quad (14)$
symmetric Mean Absolute Percentage Error (sMAPE)	$sMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\lvert A_i - F_i \rvert}{\lvert A_i \rvert + \lvert F_i \rvert} * 100\% \quad (15)$
MAD/Mean Ration (MMR)	$MMR = \frac{\sum_{i=1}^{n} \lvert A_i - F_i \rvert}{\sum_{i=1}^{n} A_i} * 100\% \quad (16)$

Initially the most common metrics for regression problems, which are Mean Absolute Error (MAE) (Function 12), and Root Mean Square Error (RMSE) (Function 13), were used. MAE, also called Mean Absolute Deviation (MAD), is the sum of the absolute errors of each observation divided by the number of observations. Mean Square Error is, respectively, the arithmetic average of the squared errors of each observation, and RMSE is simply the square root of MSE. Their definition makes it clear that the smaller these metrics are, the smaller the error is, and the better the forecast is. Those metrics must be compared to the dataset's mean and standard deviation in order to have qualitative importance.

Additionally, percentage error metrics are used, namely the Mean Absolute Percentage Error (MAPE) (Function 14), the symmetric Mean Absolute Percentage Error (sMAPE) (Function 15) and the MAD/Mean Ratio (MMR) (Function 16). In MAPE the error for each observation is divided by the actual value and then, their sum divided by the number of observations. In sMAPE the absolute difference of the predicted and actual values is divided by the absolute sum of their values. Specifically, for MAPE it should be emphasized that it is not suitable for the evaluation of the forecasts of the specific data sets, as it does not handle zero values properly, which constitute a large percentage of the datasets. Nevertheless, it is very often used in energy forecasting

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

applications, so a reference is made here as well, although it is not the primary metric for evaluating the results. Finally, MMR, also called weighted MAPE in literature, is the MAD metric divided by the arithmetical average of the observations. It is proven in many situations to be more robust, and thus more useful, than regular MAPE. [49]

6.2 Evaluation results

In this section, the preliminary results of IANOS Forecasting Engine are going to be presented. For each submodule, a table with evaluation results in terms of forecasting accuracy is going to be presented together with the informative plot for the results of the best forecasters in each situation. In each table, the metrics of the algorithms performing the best are highlighted by bold.

6.2.1 Load forecasting results

For the load forecasting sub-module, forecasts were made for the entire next day corresponding to 96 time-steps ahead, along with short-term forecasts (4 steps ahead). The size of the test set was set to 5000 time-steps, which represents the 25.96% of the dataset (19257 time-steps).

6.2.1.1 Short-term load forecasting

For short-term load forecasting the best were achieved by the Random Forest Regressor, with Light Gradient Boosting Machine being a close second, while providing the fastest results.

Table 6.2: Short-term load forecasting results

Prediction Model	MAE (W)	RMSE (W)	MMR (%)	MAPE (%)	sMAPE (%)	Time (s)
Random Forest (RF)	**163.23**	296.41	**26.51**	**27.62**	**12.04**	162.83
Gradient Boosting Regressor (GBR)	171.48	301.14	27.84	30.34	12.52	255.14
eXtreme Gradient Boosting (XGBoost)	185.39	314.65	30.1	32.72	13.48	37.45
Light Gradient Boosting Machine (LightGBM)	165.88	**295.56**	26.94	28.27	12.16	**22.59**
Support Vector Regressor with RBF kernel (RBF SVR)	172.34	323.98	27.98	27.66	12.97	159.95
Multilayer Perceptron (MLP)	206.61	363.68	33.95	37.03	16.68	42.87
Long short-term memory RNN (LSTM RNN)	167.39	322.18	27.58	26.59	12.69	53.29

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

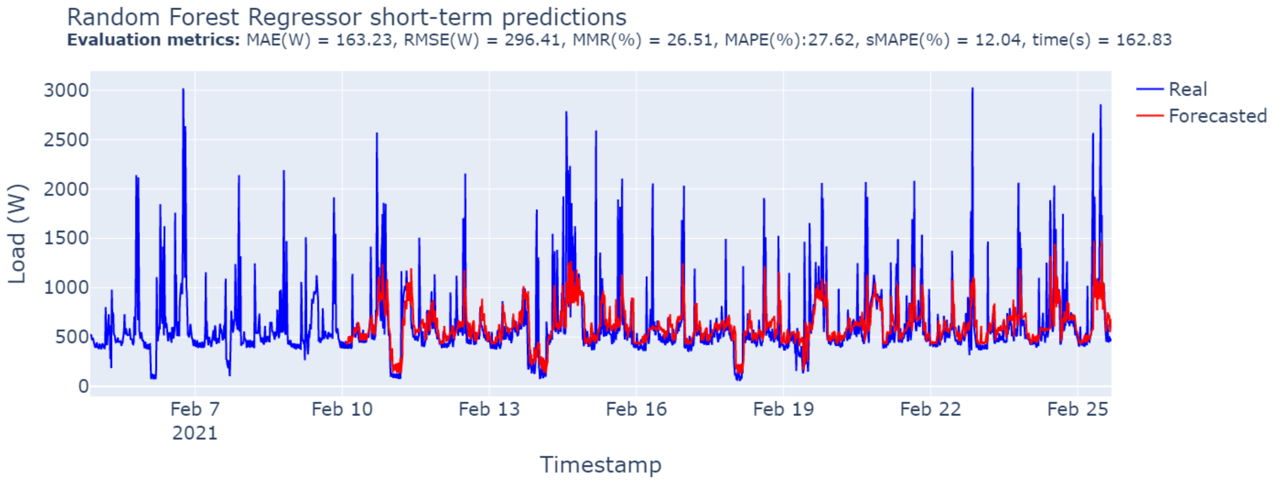

Figure 6.1: Random Forecast Regressor results for short-term load forecasting.

6.2.1.2 Long-term load forecasting

For long-term load forecasting, based on the evaluation metrics, the best were achieved by the Random Forest Regressor and Light Gradient Boosting Machine, the latter of which provided the fastest results.

Table 6.3: Long-term load forecasting results

Prediction Model	MAE (W)	RMSE (W)	MMR (%)	MAPE (%)	sMAPE (%)	Time (s)
Random Forest (RF)	**203.52**	341.76	**33.05**	**39.44**	**14.11**	3829.2
Gradient Boosting Regressor (GBR)	212.45	344.8	34.5	42.01	15.8	5664.35
eXtreme Gradient Boosting (XGBoost)	232.51	362.18	37.75	48.18	17.38	739.96
Light Gradient Boosting Machine (LightGBM)	203.68	**340.09**	**33.07**	39.86	**14.12**	**505.17**
Support Vector Regressor with RBF kernel (RBF SVR)	204.72	365.34	35.59	34.54	15.42	2863.77
Multilayer Perceptron (MLP)	213.77	376.5	35.16	40.1	16.83	1056.92

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Long short-term memory RNN (LSTM RNN)	201.24	370.07	33.1	35.53	15.75	951.37

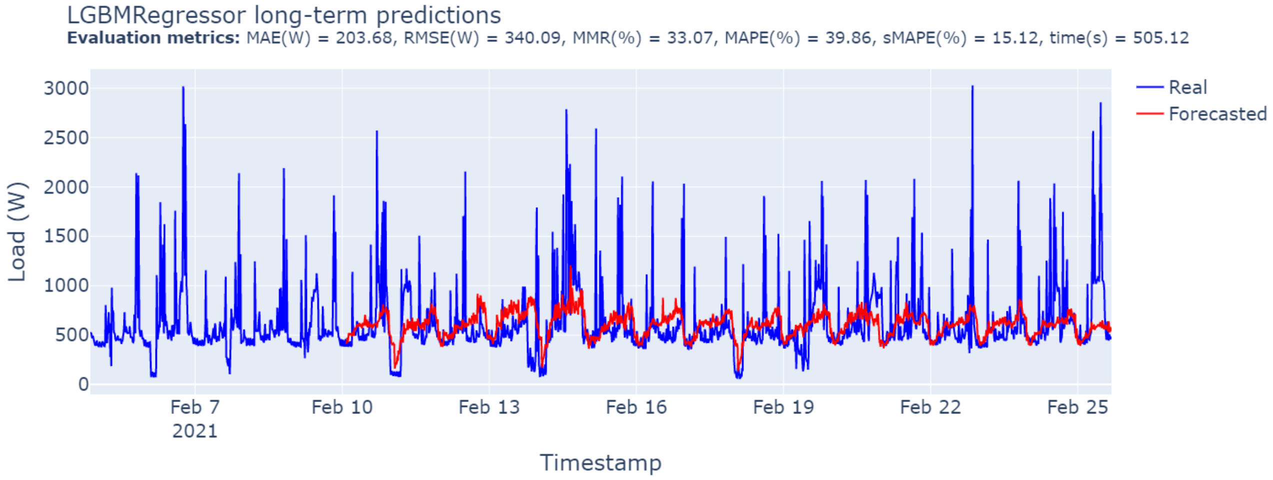

Figure 6.2: LGBMRegressor results for long-term load forecasting.

As observed by the plots the results of the residential load forecasting are not satisfactory, this is reasonable due to the highly unpredictable nature of the residential consumption profiles and low historical data availability. In the following version of the deliverable, the highest amount data that are going to be available will assist in delivering results that are more accurate. Based again on the plots, XGBoost Regressor could be considered a more appropriate predictor for existing data.

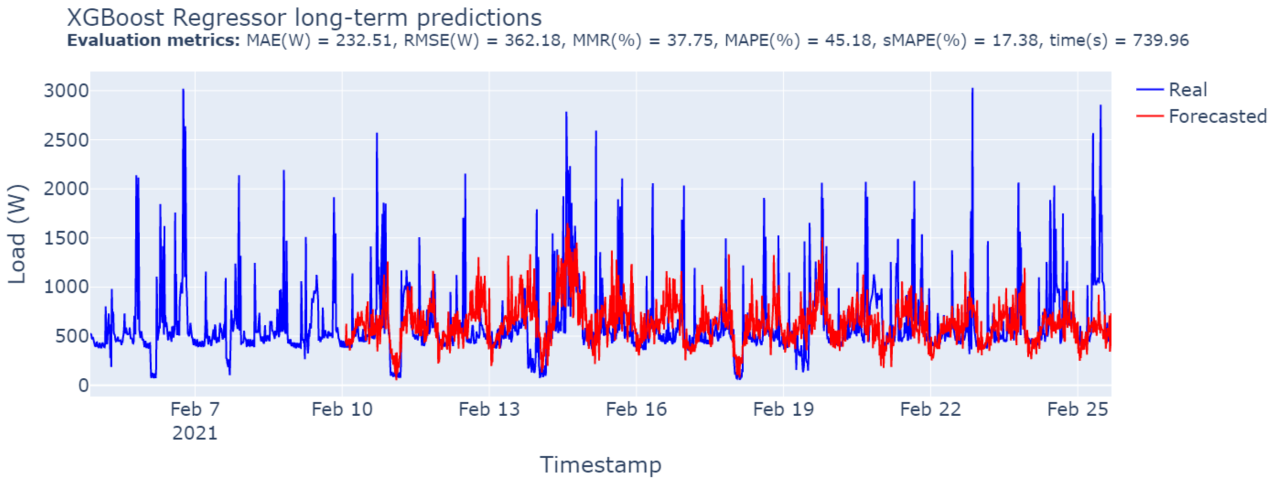

Figure 6.3: XGBoost results for long-term load forecasting.

6.2.2 PV generation forecasting results

With the same logic as in the load forecasting, for the PV generation forecasting sub-module day ahead forecasts that correspond to 96 time-steps ahead, together with short-term predictions (4 steps ahead) were held. The size of the test set was set to 5000 time-steps, which represents the 22.36% of the dataset (22364 time-steps).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

6.2.2.1 Short-term PV generation forecasting

For short-term PV generation forecasting the best were achieved by Light Gradient Boosting Machine, which also provided the fastest results.

Table 6.4: Short-term PV generation forecasting results

Prediction Model	MAE (kW)	RMSE (kW)	MMR (%)	sMAPE (%)	Time (s)
Random Forest (RF)	275.68	472.79	22.33	44.15	116.53
Gradient Boosting Regressor (GBR)	357.15	555.06	28.92	45.58	147.93
eXtreme Gradient Boosting (XGBoost)	397.0	622.73	32.15	42.1	35.6
Light Gradient Boosting Machine (LightGBM)	**271.72**	**462.48**	**22.01**	**41.61**	**23.29**
Support Vector Regressor with RBF kernel (RBF SVR)	275.81	459.48	22.34	41.68	93.84
Multilayer Perceptron (MLP)	1184.9	1646.01	95.96	71.89	48.58
Long short-term memory RNN (LSTM RNN)	287.83	492.93	23.31	48.25	48.41

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

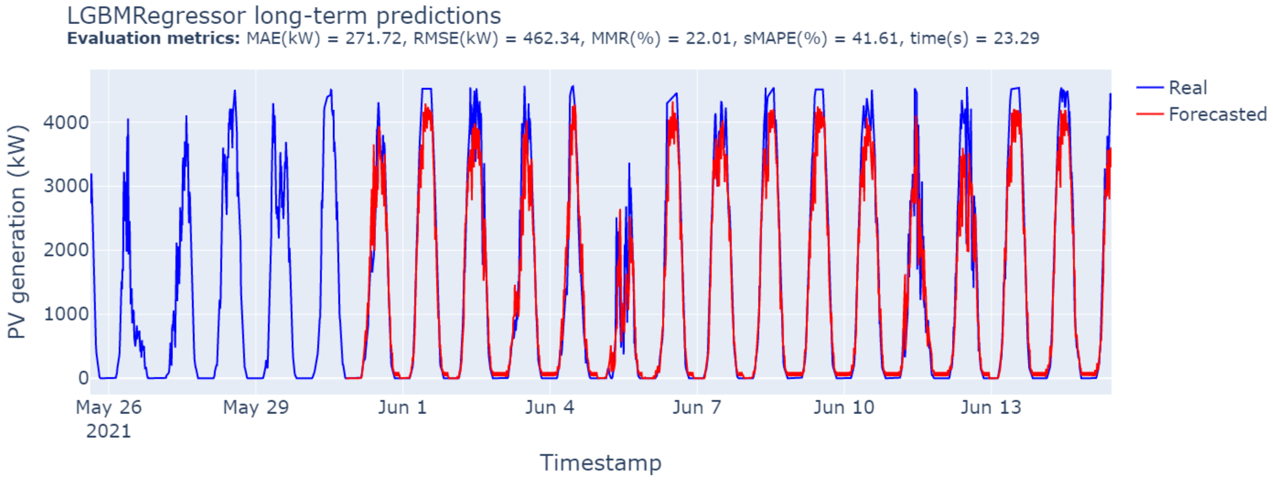

Figure 6.4: LGBMRegressor results for short-term PV generation forecasting.

6.2.2.2 Long-term PV generation forecasting

For long-term PV generation forecasting the best were achieved by Support Vector Regressor (SVR), with Light Gradient Boosting Machine providing decent and the faster results.

Table 6.5: Long-term PV generation forecasting results

Prediction Model	MAE (kW)	RMSE (kW)	MMR (%)	sMAPE (%)	Time (s)
Random Forest (RF)	449.12	731.45	36.37	50.26	2526.15
Gradient Boosting Regressor (GBR)	521.53	817.27	42.24	50.1	3720.78
eXtreme Gradient Boosting (XGBoost)	564.91	931.25	45.75	45.28	703.86
Light Gradient Boosting Machine (LightGBM)	438.09	753.9	35.48	44.56	555.20
Support Vector Regressor with RBF kernel (RBF SVR)	**425.33**	**717.55**	**34.45**	**45.02**	**2401.57**
Multilayer Perceptron (MLP)	1009.51	1388.44	81.75	67.77	897.16
Long short-term memory RNN (LSTM RNN)	437.52	715.3	35.43	55.54	1364.83

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

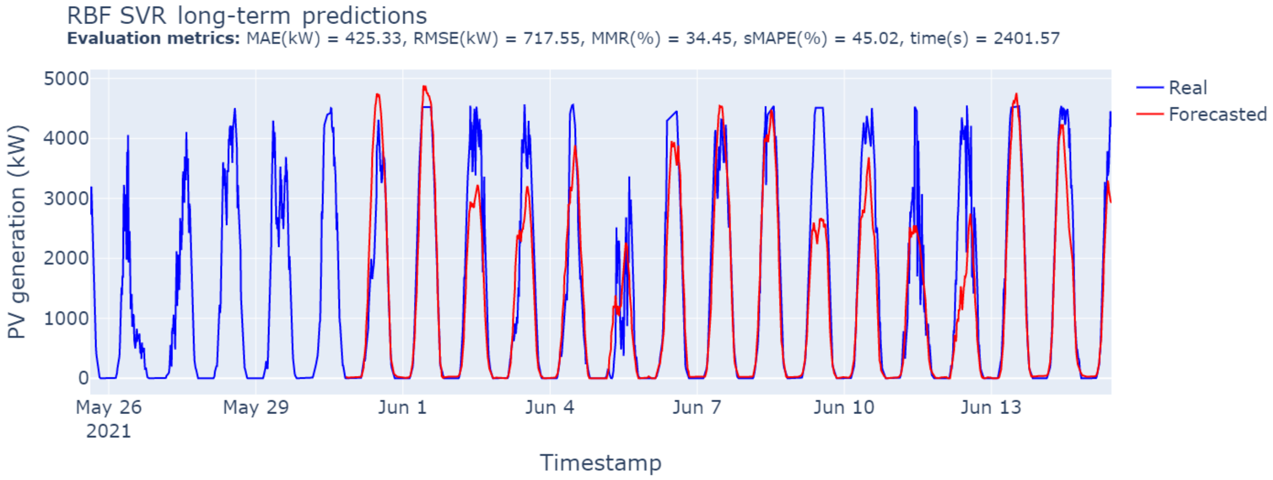

Figure 6.5: RBF SVR results for long-term PV generation forecasting.

6.2.3 Wind generation forecasting results

Same as in the PV generation forecasting, the wind generation forecasting sub-module provide day ahead forecasts that correspond to 24 time-steps ahead, together with short-term predictions (1 step ahead). The size of the test set was set to 3500 time-steps, which represents the 20.88% of the dataset (16765 time-steps).

6.2.3.1 Short-term wind generation forecasting

For short-term wind generation forecasting the best were achieved by Light Gradient Boosting Machine, which also provided the fastest results.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Table 6.6: Short-term wind generation forecasting results

Prediction Model	MAE	RMSE	MMR (%)	sMAPE (%)	Time (s)
Random Forest (RF)	0.07	0.11	18.49	20.35	11.29
Gradient Boosting Regressor (GBR)	0.07	0.11	18.85	20.28	15.49
eXtreme Gradient Boosting (XGBoost)	0.07	0.11	19.65	21.17	4.58
Light Gradient Boosting Machine (LightGBM)	**0.07**	**0.1**	**18.38**	**20.16**	**3.92**
Support Vector Regressor with RBF kernel (RBF SVR)	0.07	0.11	19.77	21.12	11.73
Multilayer Perceptron (MLP)	0.14	0.19	40.66	33.87	8.73
Long short-term memory RNN (LSTM RNN)	0.07	0.11	20.53	24.54	13.55

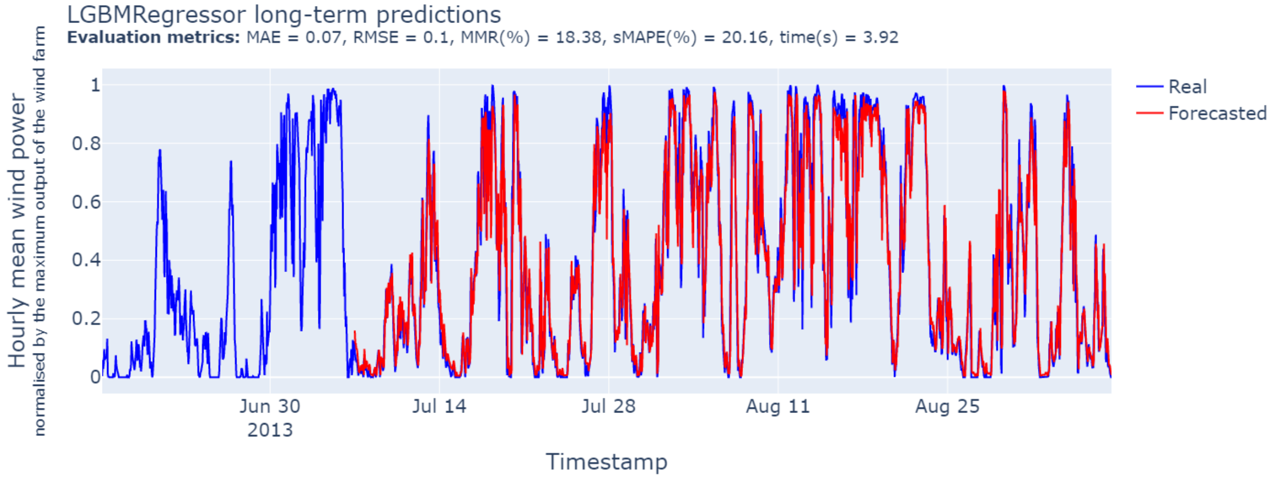

Figure 6.6: LGBMRegressor results for short-term wind generation forecasting.

6.2.3.2 Long-term wind generation forecasting

Same as in short-term, in long-term wind generation forecasting the best were achieved by Light Gradient Boosting Machine, which also provided the fastest results.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

Table 6.7: Long-term wind generation forecasting results

Prediction Model	MAE	RMSE	MMR (%)	sMAPE (%)	Time (s)
Random Forest (RF)	0.14	0.18	36.37	29.27	273.12
Gradient Boosting Regressor (GBR)	0.14	0.18	35.87	29.29	368.7
eXtreme Gradient Boosting (XGBoost)	0.14	0.19	36.95	30.59	110.43
Light Gradient Boosting Machine (LightGBM)	**0.13**	**0.18**	**34.75**	**28.53**	**93.64**
Support Vector Regressor with RBF kernel (RBF SVR)	0.2	0.26	53.22	37.64	314.19
Multilayer Perceptron (MLP)	0.19	0.26	55.8	41.4	225.95
Long short-term memory RNN (LSTM RNN)	0.16	0.23	47.2	39.31	448.59

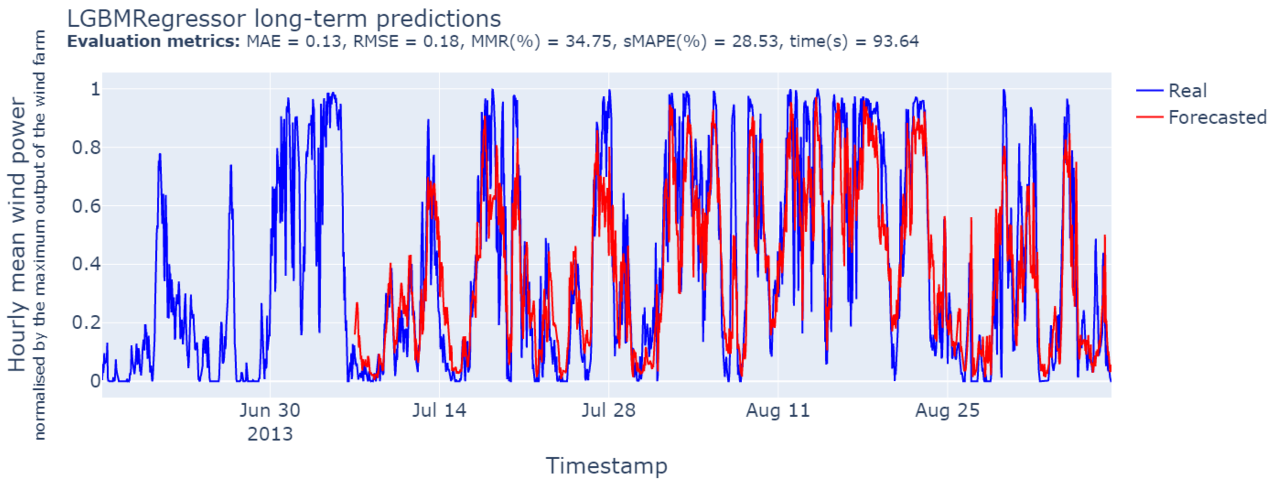

Figure 6.7: LGBMRegressor results for long-term wind generation forecasting.

6.2.4 Energy market forecasting results

The energy market forecasts were separated into four different forecasts each one with its separate forecast horizon.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

6.2.4.1 Day-ahead price

For the day-ahead price forecast the horizon is set to day-ahead (24 time-steps). The size of the test set was set to 1500 time-steps, which represents the 16.89% of the dataset (8880 time-steps). The best results were achieved by the Support Vector Regressor with RBF kernel.

Table 6.8: Day-ahead price forecasting results

Prediction Model	MAE (EUR/kWh)	RMSE (EUR/kWh)	MMR (%)	MAPE (%)	sMAPE (%)	Time (s)
Random Forest (RF)	0.01	0.02	15.84	339.15	9.67	148.51
Gradient Boosting Regressor (GBR)	0.01	0.02	17.02	321.38	10.4	195.25
eXtreme Gradient Boosting (XGBoost)	0.01	0.02	16.07	229.49	9.76	64.41
Light Gradient Boosting Machine (LightGBM)	0.01	0.02	15.83	308.54	9.52	**53.49**
Support Vector Regressor with RBF kernel (RBF SVR)	**0.01**	**0.01**	**11.46**	**377.7**	**6.96**	112.32
Multilayer Perceptron (MLP)	0.02	0.02	20.93	414.54	11.64	119.63
Long short-term memory RNN (LSTM RNN)	0.01	0.02	26.87	740.88	14.63	172.11

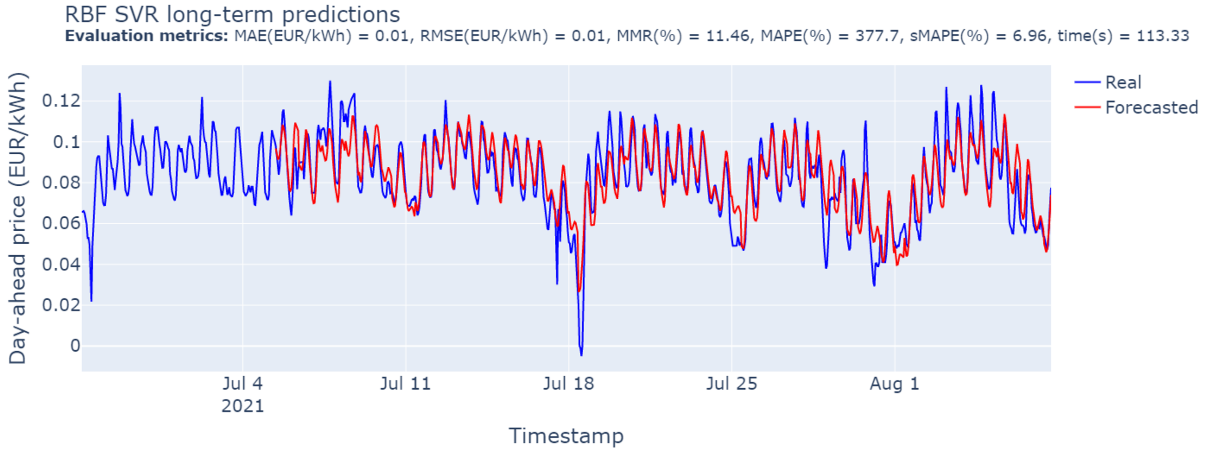

Figure 6.8: RBF SVR results for day-ahead price forecasting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

6.2.4.2 Intra-day price

For the intra-day price forecast, the horizon is set to next hour (2 time-steps). The size of the test set was set to 2000 time-steps, which represents the 25.25% of the dataset (7920 time-steps). The best results were achieved by the Random Forest Regressor, with Light Gradient Boosting Machine being the second best and fastest algorithm.

Table 6.9: Intra-day price forecasting results

Prediction Model	MAE (EUR/kWh)	RMSE (EUR/kWh)	MMR (%)	sMAPE (%)	Time (s)
Random Forest (RF)	**0.01**	**0.01**	**9.54**	**5.97**	15.09
Gradient Boosting Regressor (GBR)	0.01	0.01	10.06	6.25	21.4
eXtreme Gradient Boosting (XGBoost)	0.01	0.01	10.55	6.52	4.69
Light Gradient Boosting Machine (LightGBM)	**0.01**	**0.01**	**9.65**	**6.0**	**3.75**
Support Vector Regressor with RBF kernel (RBF SVR)	0.01	0.01	11.85	7.13	7.25
Multilayer Perceptron (MLP)	0.02	0.03	34.01	20.6	7.83
Long short-term memory RNN (LSTM RNN)	0.01	0.01	15.03	8.46	11.72

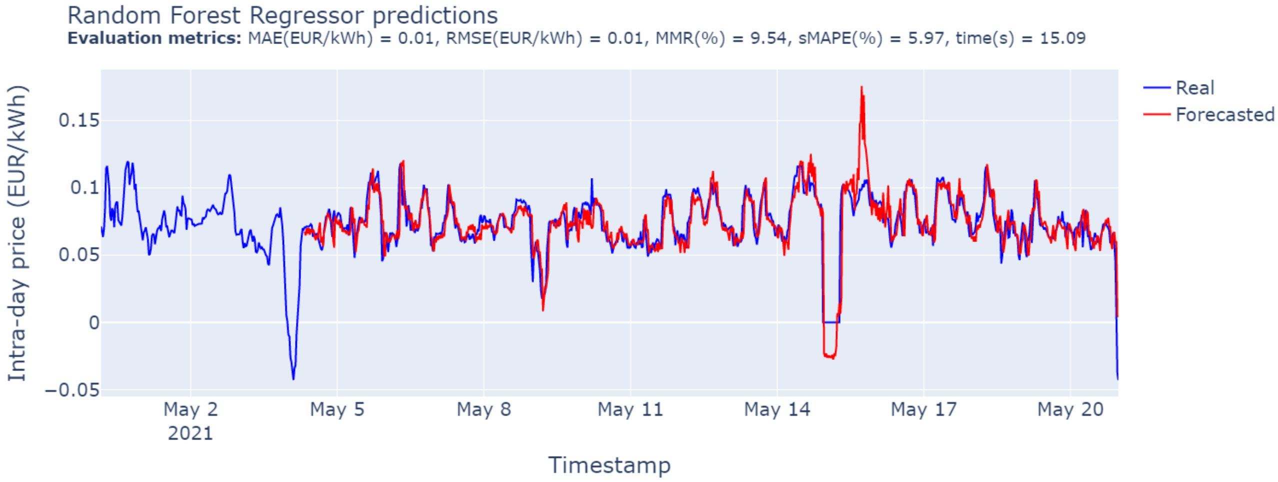

Figure 6.9: Random Forest Regressor results for intra-day price forecasting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

6.2.4.3 Imbalance price

For the imbalance price forecast, the horizon is set to one quarter ahead (1 timestep). The size of the test set was set to 7000 time-steps, which represents the 19.71% of the dataset (35516 time-steps). The best results were achieved by the RBF SVR, with Light Gradient Boosting Machine being the second best and fastest algorithm.

Table 6.10: Imbalance price forecasting results

Prediction Model	MAE (EUR/kWh)	RMSE (EUR/kWh)	MMR (%)	sMAPE (%)	Time (s)
Random Forest (RF)	0.03	0.04	45.39	28.26	26.29
Gradient Boosting Regressor (GBR)	0.04	0.04	46.62	28.82	34.15
eXtreme Gradient Boosting (XGBoost)	0.04	0.05	47.83	29.34	10.21
Light Gradient Boosting Machine (LightGBM)	**0.03**	**0.04**	**45.47**	**28.28**	**9.17**
Support Vector Regressor with RBF kernel (RBF SVR)	**0.03**	**0.04**	**39.74**	**24.61**	59.15
Multilayer Perceptron (MLP)	0.04	0.05	58.57	34.08	15.5
Long short-term memory RNN (LSTM RNN)	0.04	0.05	47.37	29.23	28.57

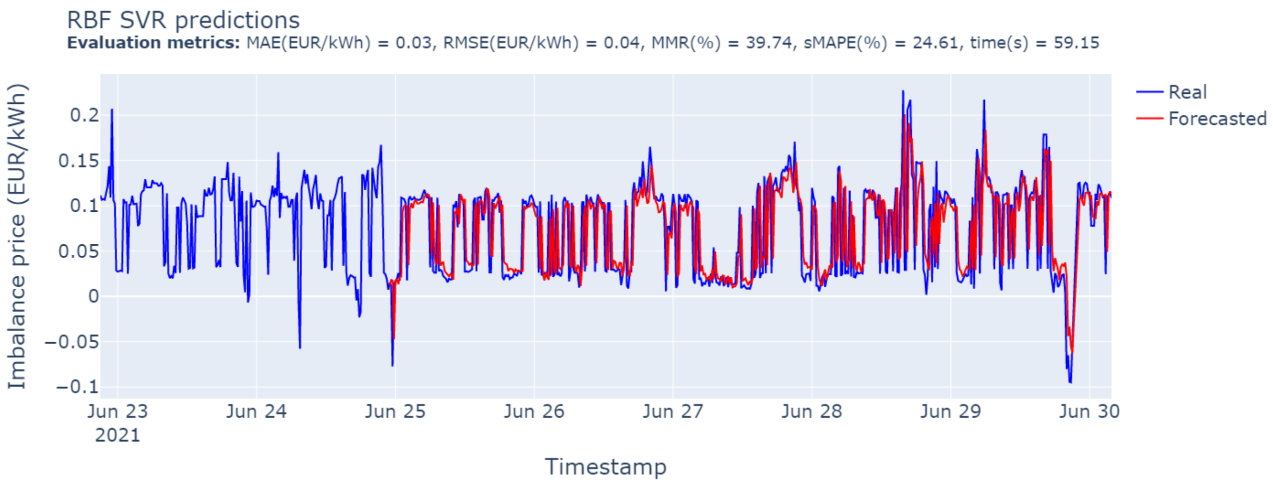

Figure 6.10: RBF SVR results for imbalance price forecasting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

6.2.4.4 FCR

For the FCR forecast, the horizon is set to next hour (1 timestep). The size of the test set was set to 500 time-steps, which represents the 22.4% of the dataset (2232 time-steps). The best results were achieved by the Random Forest Regressor, with Light Gradient Boosting Machine being the second best and fastest algorithm.

Table 6.11: FCR forecasting results

Prediction Model	MAE (EUR/MW/ISP)	RMSE (EUR/MW/ISP)	MMR (%)	MAPE (%)	sMAPE (%)	Time (s)
Random Forest (RF)	**0.32**	**0.62**	**7.23**	**6.76**	**3.32**	1.22
Gradient Boosting Regressor (GBR)	0.36	0.7	8.08	7.27	3.58	1.37
eXtreme Gradient Boosting (XGBoost)	0.34	0.65	7.7	6.94	3.46	0.630
Light Gradient Boosting Machine (LightGBM)	**0.33**	**0.59**	**7.41**	**6.8**	**3.4**	**0.55**
Support Vector Regressor with RBF kernel (RBF SVR)	0.44	0.74	9.95	8.66	4.47	0.72
Multilayer Perceptron (MLP)	0.81	1.05	20.3	21.8	10.43	1.1
Long short-term memory RNN (LSTM RNN)	0.34	0.64	33.83	34.24	16.22	5.17

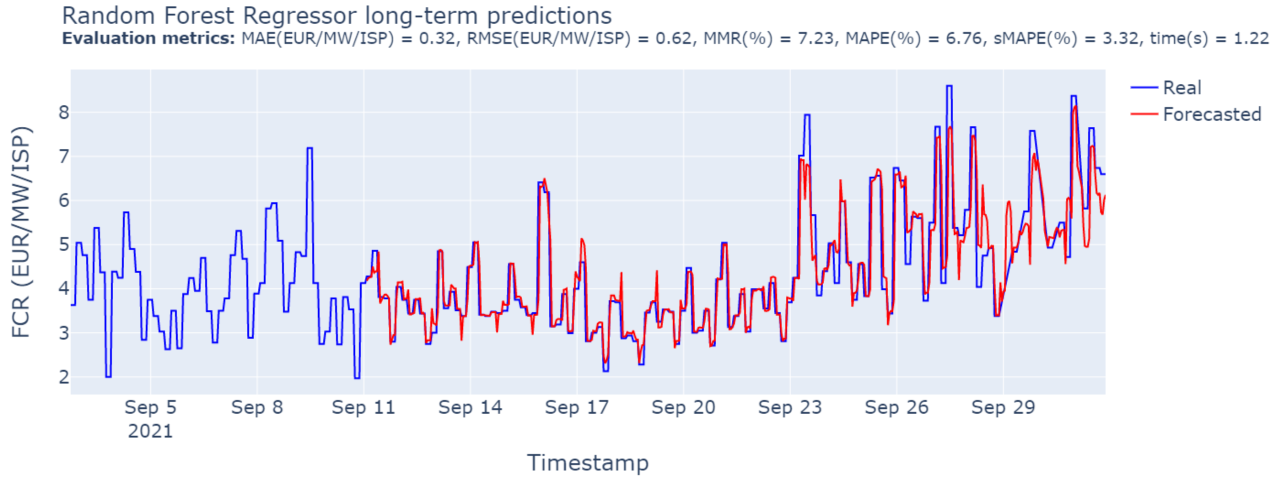

Figure 6.11: Random Forest Regressor results for FCR forecasting.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

7 Conclusions and future steps

The presented deliverable provides the 1st version of IANOS iVPP Forecasting Engine component and its functionalities. As highlighted in the document, the component is separated into different subcomponents, namely load, generation (PV and wind) and energy market forecasting. The component in its current implementation utilizes data-driven models, based on the most used machine learning algorithms to provide the necessary forecasts. In most of the forecasts, despite the lack of available data in many cases, the forecasting results can be characterized as accurate, compared to the commonly used techniques for energy load, generation and market forecasting.

More specifically, for load forecasting the most appropriate models are Random Forest Regressor for short-term and LGBMRegressor for long-term forecasting, for PV generation forecasting LGBMRegressor for short-term and RBF SVR for long-term forecasting, for wind generation forecasting LGBMRegressor for both short- and long-term forecasting, for energy market forecasting and RBF SVR was the best performing model for day-ahead and imbalance price forecasting, while Random Forest Regressor performed best for intra-day price and FCR forecasting. It is possible that these results will change in the second version of the deliverable, where there will be more data available.

The pending actions include the collection of all the necessary data from IANOS pilots along with integration with the ESB component to achieve full-scale operation, the inclusion of the physical models in the Forecasting Engine, further exploration of machine learning techniques for energy related time series forecasting, and finally the expansion of the Forecasting Engine for thermal load and thermal PV generation forecasting. The second and final version of the iVPP Forecasting Engine will be described in D4.4, which will be delivered in M30 of IANOS project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 957810

8 References

[1] A. Lahouar and J. ben Hadj Slama, “Random forests model for one day ahead load forecasting,” Nov. 2015. doi: 10.1109/IREC.2015.7110975.

[2] N. Fumo and M. A. Rafe Biswas, “Regression analysis for prediction of residential energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 47. Elsevier Ltd, pp. 332–343, 2015. doi: 10.1016/j.rser.2015.03.035.

[3] E. M. Sobrino, A. V. Santiago, and A. M. González, “Forecasting the Electricity Hourly Consumption of Residential Consumers with Smart Meters using Machine Learning Algorithms,” in *2019 IEEE Milan PowerTech*, 2019, pp. 1–6. doi: 10.1109/PTC.2019.8810902.

[4] P. Amin, L. Cherkasova, R. Aitken, and V. Kache, “Analysis and Demand Forecasting of Residential Energy Consumption at Multiple Time Scales,” in *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 2019, pp. 494–499.

[5] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, “A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting,” *Energy*, vol. 214, Jan. 2021, doi: 10.1016/j.energy.2020.118874.

[6] C. Tian, J. Ma, C. Zhang, and P. Zhan, “A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network,” *Energies*, vol. 11, no. 12, Dec. 2018, doi: 10.3390/en11123493.

[7] entsoe.eu, “ENTSO-E Transparency Platform.” https://www.entsoe.eu/data/ (accessed Nov. 15, 2021).

[8] R. Rajabi and A. Estebsari, “Deep Learning Based Forecasting of Individual Residential Loads Using Recurrence Plots,” 2019.

[9] A. Almalaq and G. Edwards, “Comparison of Recursive and Non-Recursive ANNs in Energy Consumption Forecasting in Buildings,” in *2019 IEEE Green Technologies Conference(GreenTech)*, 2019, pp. 1–5. doi: 10.1109/GreenTech.2019.8767130.

[10] T. Ploysuwan, “Deep CNN amp; LSTM network for appliances energy forecasting in residential houses using IoT sensors,” in *2019 7th International Electrical Engineering Congress (iEECON)*, 2019, pp. 1–4. doi: 10.1109/iEECON45304.2019.8938914.

[11] N. Shabbir, R. Amadiahangar, H. A. Raja, L. Kütt, and A. Rosin, “Residential Load Forecasting Using Recurrent Neural Networks,” in *2020 IEEE 14th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, 2020, vol. 1, pp. 478–481. doi: 10.1109/CPE-POWERENG48600.2020.9161565.

[12] A. Dolara, S. Leva, and G. Manzolini, “Comparison of different physical models for PV power output prediction,” *Solar Energy*, vol. 119, pp. 83–99, 2015, doi: https://doi.org/10.1016/j.solener.2015.06.017.

[13] W. F. Holmgren, A. T. Lorenzo, and C. Hansen, “A Comparison of PV Power Forecasts Using PVLib-Python,” in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*, 2017, pp. 1127–1131. doi: 10.1109/PVSC.2017.8366724.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 957810

[14] J. Zeng and W. Qiao, “Short-term solar power prediction using a support vector machine,” *Renewable Energy*, vol. 52, pp. 118–127, 2013, doi: https://doi.org/10.1016/j.renene.2012.10.009.

[15] J. Huang, A. Troccoli, and P. Coppin, “An analytical comparison of four approaches to modelling the daily variability of solar irradiance using meteorological records,” *Renewable Energy*, vol. 72, pp. 195–202, 2014, doi: https://doi.org/10.1016/j.renene.2014.07.015.

[16] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, “Day-Ahead Power Output Forecasting for Small-Scale Solar Photovoltaic Electricity Generators,” *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2253–2262, 2015, doi: 10.1109/TSG.2015.2397003.

[17] J.-I. Lee, I.-W. Lee, and S.-H. Kim, “Multi-site photovoltaic power generation forecasts based on deep-learning algorithm,” in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 2017, pp. 1118–1120. doi: 10.1109/ICTC.2017.8190872.

[18] W. Lee, K. Kim, J. Park, J. Kim, and Y. Kim, “Forecasting Solar Power Using Long-Short Term Memory and Convolutional Neural Networks,” *IEEE Access*, vol. 6, pp. 73068–73080, 2018, doi: 10.1109/ACCESS.2018.2883330.

[19] K. Wang, X. Qi, and H. Liu, “A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network,” *Applied Energy*, vol. 251, p. 113315, 2019, doi: https://doi.org/10.1016/j.apenergy.2019.113315.

[20] D. Su, E. Batzelis, and B. Pal, “Machine Learning Algorithms in Forecasting of Photovoltaic Power Generation,” in *2019 International Conference on Smart Energy Systems and Technologies (SEST)*, 2019, pp. 1–6. doi: 10.1109/SEST.2019.8849106.

[21] S. Leva, M. Mussetta, A. Nespoli, and E. Ogliari, “PV power forecasting improvement by means of a selective ensemble approach,” Nov. 2019, pp. 1–5. doi: 10.1109/PTC.2019.8810921.

[22] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, “Current methods and advances in forecasting of wind power generation,” *Renewable Energy*, vol. 37, no. 1, pp. 1–8, 2012, doi: https://doi.org/10.1016/j.renene.2011.05.033.

[23] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, “Long-term wind speed and power forecasting using local recurrent neural network models,” *IEEE Transactions on Energy Conversion*, vol. 21, no. 1, pp. 273–284, 2006, doi: 10.1109/TEC.2005.847954.

[24] R. Jursa and K. Rohrig, “Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models,” *International Journal of Forecasting*, vol. 24, pp. 694–709, Nov. 2008, doi: 10.1016/j.ijforecast.2008.08.007.

[25] R. Jursa, “Wind power prediction with different artificial intelligence models,” Nov. 2010.

[26] J. Chen, G.-Q. Zeng, W. Zhou, W. Du, and K.-D. Lu, “Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization,” *Energy Conversion and Management*, vol. 165, pp. 681–695, 2018, doi: https://doi.org/10.1016/j.enconman.2018.03.098.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 957810

[27] K. Maciejowska, W. Nitka, and T. Weron, “Day-Ahead vs. Intraday—Forecasting the Price Spread to Maximize Economic Benefits,” *Energies*, vol. 12, no. 4, 2019, doi: 10.3390/en12040631.

[28] W. Sun, J.-C. Lu, and M. Meng, “APPLICATION OF TIME SERIES BASED SVM MODEL ON NEXT-DAY ELECTRICITY PRICE FORECASTING UNDER DEREGULATED POWER MARKET,” 2006.

[29] E. E. Elattar, “Day-ahead price forecasting of electricity markets based on local informative vector machine,” *IET Generation, Transmission and Distribution*, vol. 7, no. 10, pp. 1063–1071, 2013, doi: 10.1049/iet-gtd.2012.0382.

[30] C. González, J. Mira-McWilliams, and I. Juárez, “Important variable assessment and electricity price forecasting based on regression tree models: Classification and regression trees, Bagging and Random Forests,” *IET Generation, Transmission and Distribution*, vol. 9, no. 11, pp. 1120–1128, Aug. 2015, doi: 10.1049/iet-gtd.2014.0655.

[31] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, “Ensemble of relevance vector machines and boosted trees for electricity price forecasting,” *Applied Energy*, vol. 250, pp. 540–548, Sep. 2019, doi: 10.1016/j.apenergy.2019.05.062.

[32] U. Ugurlu, I. Oksuz, and O. Tas, “Electricity price forecasting using recurrent neural networks,” *Energies*, vol. 11, no. 5, 2018, doi: 10.3390/en11051255.

[33] S. Anbazhagan and N. Kumarappan, “Day-ahead deregulated electricity market price forecasting using recurrent neural network,” *IEEE Systems Journal*, vol. 7, no. 4, pp. 866–872, 2013, doi: 10.1109/JSYST.2012.2225733.

[34] E. Kraft, J. Rominger, V. Mohiuddin, and D. Keles, “Forecasting of Frequency Containment Reserve Prices using Econometric and Artificial Intelligence Approaches.” [Online]. Available: www.fzi.de

[35] “Python,” *https://www.python.org/*. https://www.python.org/ (accessed Dec. 22, 2021).

[36] “Numpy.” http://www.numpy.org/ (accessed Dec. 22, 2021).

[37] “Pandas.” https://pandas.pydata.org/ (accessed Dec. 22, 2021).

[38] “Scikit-learn.” https://scikit-learn.org (accessed Dec. 22, 2021).

[39] “Keras.” https://keras.io/ (accessed Dec. 22, 2021).

[40] “Plotly.” https://plotly.com/ (accessed Dec. 22, 2021).

[41] statisticshowto.com, “Z-Score: Definition, Formula and Calculation.” https://www.statisticshowto.com/probability-and-statistics/z-score/ (accessed Nov. 09, 2021).

[42] scikit-learn.org, “sklearn.preprocessing.MinMaxScaler.” https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html (accessed Nov. 29, 2021).

[43] scikit-learn.org, “sklearn.preprocessing.StandardScaler.” https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (accessed Nov. 11, 2021).

[44] W. Holmgren, C. Hansen, and M. Mikofski, “pvlib python: a python package for modeling solar energy systems,” *Journal of Open Source Software*, vol. 3, p. 884, Nov. 2018, doi: 10.21105/joss.00884.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 957810

[45] AEC, “ZONNEPARK AMELAND.” http://www.amelandenergie.nl/files/dashboard.html (accessed Nov. 30, 2021).

[46] Robert Gasch, *Windkraftanlagen*. 2010.

[47] Natasha Sharma, “Wind Forecasting Dataset.” https://www.kaggle.com/natasha23/wind-forecasting-dataset/metadata (accessed Feb. 08, 2022).

[48] “bmrs ELEXON.” https://www.bmreports.com/bmrs/?q=eds/main (accessed Jan. 05, 2022).

[49] S. Kolassa and W. Schütz, “Advantages of the MAD/mean ratio over the MAPE,” *Foresight: The International Journal of Applied Forecasting*, vol. 6, pp. 40–43, Nov. 2007.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 957810