TRUSTAI

TRANSPARENT, RELIABLE
& UNBIASED SMART TOOL

Use Case 3 –Energy

D7.2 - Initial validation of the explainable AI models from business experts

APINTECH

September, 30, 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No 952060.

DOCUMENT CONTROL PAGE

DOCUMENT	D7.2 – Initial validation of the explainable AI models from business experts
TYPE	Report
DISTRIBUTION LEVEL	Public
DUE DELIVERY DATE	30/09/2022
DATE OF DELIVERY	30/09/2022
VERSION	FINAL
DELIVERABLE RESPONSIBLE	APINTECH
AUTHOR (S)	Shah, Merkos, Leite, Daskalalis, Sakkas Nikos, Sakkas Nikitas
OFFICIAL REVIEWER/s	Leite, Barbu

DOCUMENT HISTORY

VERSION	AUTHORS	DATE	CONTENT AND CHANGES
0.1	Shah, Merkos, Leite, Daskalalis, Sakkas Nikos, Sakkas Nikitas	31.07	First draft
0.2	Leite, Barbu	28.08	Review
0.3	Sakkas	06.09	Final version
1	APINTECH	17.01.23	Revision following EC comments and requests

DISCLAIMER:

Executive Summary

This deliverable is based on extensive data sets and modelling work carried out thereupon, and seeks to introduce the concepts of explainability in the energy domain in two different scopes and subcases; that of the country level forecasts as well as, more importantly, that of the building level ones.

Explainability considerations are indeed in their infancy in the energy literature. Very recently (2022) the issue has started receiving some attention in the building related literature, while at the country level, we believe we are the first ever to raise the issue in the literature (2021).

The building energy forecasting literature is predominantly based on neural network (NN) approaches. Paradoxically, the key approach is to extract explanations from such models. The idea of looking at genetic programming (GP) approaches and exploring their potential to provide for explainability has just not yet received any attention.

Thus, the challenge addressed in this deliverable as regards the building subcase is to benchmark the performance of dominant NN models against emerging GP approaches. This benchmark needs to cover both accuracy as well as explainability issues. Real time data from the BMS of an office building has been used as a data source for this extensive benchmarking exercise. Metrics to be used for the benchmarking are the typical ones used for model accuracy evaluation: MAE, RMSE and MAPE

At the time of completion of this work no conclusive evidence has been yet reached. In fact, although the accuracy of the GP approaches seems well comparable with the NN approaches, it remains to be seen whether, especially the symbolic expressions that can be derived thereof can be compact enough to be used as basis for explanation. Of course it remains to be seen what the complexity of these equations will be and whether they can provide for user friendly explanations. However this comes later in the agenda. At this point, the key issue is the benchmarking issue.

As regards the country level subcase we have practically introduced local explanation concepts in terms of counterfactual analysis. Some interesting results have been reached and have been published, shedding light on especially controversial issues in the literature, such as the demand elasticity of fuel prices/ taxes.

Industry stakeholder inputs have been sought and taken account of in the roll out of the work. These issues are partially also reported in D 2.1. However, this deliverable presents a more comprehensive overview of the multifold expert validation activities and the insights that have been achieved through this important, ongoing and intensifying activity.

In short, topics such as counterfactual analysis and feature importance were highlighted as to their importance. Symbolic expressions were not discussed with the stakeholders in any significant detail. The main reason is that GP approaches are totally outside the current state of the art so no true feedback could possibly result.

In the coming period the challenges are as follows:

- As regards the building subcase,
    - We look forward to exploring the TRUST AI framework that will allow us to prune the resulting GP expressions in search of compact and meaningful expressions that can provide better insights on building performance as well as several side issues (seasonality differences) and the ability of users to interact with them in a rational, energy efficiency-wise way.
    - We look forward to integrating these developments within a commercial venture that is unfolding at this moment in time, outside TRUST AI, which however can serve as a practical testbed of making use of the TRUST AI framework services, and making use of the GP and other models customised within it.

- As regards the country subcase
    - We look forward to introducing GP approaches to see how they perform vis a vis the dominant NN ones again in terms of comparative accuracies as well as explainability potential of the underlying symbolic expressions.
    - We look forward to exploring the concept of so-called *ensemble counterfactuals*, whereby all three consumption areas (residences, transport and industry) will be brought together to elaborate local explanation concepts and related fuel price and tax elasticity.

Table of Contents

Abbreviations and Acronyms

AI	Artificial Intelligence
EC	European Commission
EU	European Union
HCXAI	Human-centred Explainable AI
KPI	Key Performance Indicators
MS	Milestones
PM	Person Month
PR	Press Release
SMEs	Small and Medium-sized Enterprises
WP	Work Package
XAI	Explainable Artificial Intelligence
GP	Genetic Programming
BMS	Building Management System
NN	Neural Network
JASP	Jeffreys's Amazing Statistics Program

1. Introduction

The deliverable presents some early trials on the Energy Use Case, in both subcases investigated, country and building level forecasting. The key aspiration of the deliverable is to **investigate the potential of explainability and GP models in particular in the energy realm and to benchmark it against traditional neural network (NN) based approaches.** This objective will underlie all future work and technical development, and will define the concrete TRUST-AI framework services, pertinent for the energy use case.

In the course of this work we have significantly expanded the initial (work-programme) scope of the energy use case. Although the building level forecasting continues to assume the predominant position in the research we have also included one more variant, the country level. Thus, the energy use case comprises all possible scopes of demand forecasting which in summary are:

1. **building level demand forecasting**, whereby the forecast is carried out one day/ one week ahead with an hourly resolution. This approach is highly pertinent also for **aggregate level forecasting**, whereby the forecast is, again, carried out one day/ one week ahead with an hourly resolution. Aggregate demand assumes more and more importance, as concepts such as energy communities, micro grids, etc., emerge and proliferate as key components of the energy transition.
2. **country level forecasting**, whereby the forecast is carried out on a yearly resolution.

The work described in [1] will be referred to as the **short term use case**, while [2] as the respective **long term use case.** Additionally, **in short term forecasting we will focus on electricity, rather than building energy at large**. There is no inherent limitation in this decision. It is driven by the fact that electricity is gaining increasing interest, as electrification is a key direction in energy transition. Additionally, electricity meters are becoming more and more available and operate at high resolutions. This is not the case for gas/ oil meters. Overall however, the electricity case can be seamlessly expanded to any building energy carrier, provided data are in place. In the long term use case, both electricity as well as final energy consumption are considered.

The deliverable builds on the specification document, deliverable D 7.1, while a number of contacts and interviews with industry stakeholders (reported in D 2.1) have provided significant orientation inputs.

The deliverable will guide all future work related to the energy use case, whereby explainability will be practically implemented in the energy demand forecasting approach.

Results achieved in this deliverable have been presented in a global energy related conference, published in a high impact factor journal and in two peer-reviewed flash papers. The references will follow in the text below.

In the text below we will present:

- a **description of the demand forecasting problem** with a focus on the decisions it can support (Section 2)
- a **summary of the literature approaches** in the area (Section 3)
- the **description of the approach pursued in all two sub-scopes/ cases** as far as data and modelling is concerned and the **results that have been reached** (Section 4)
- a **first concept of explainability, for all two sub-scopes**, based on the above results as well as industry inputs (Section 5)
- the **plan for the future development of the use case and the potential for uptake in major, related, product development** that is carried in at this moment within the POLIS-21 group. Although this technology development is not within the scope of TRUST-AI it nevertheless offers an important opportunity to practically up-take explainability concepts and illustrate in a tangible way the added value that can result thereof. Additionally, the time plan is very much aligned with that of TRUST-AI which makes this uptake fully realistic (Section 6).

2. Problem description. Linking demand forecasting to decisions

Below, we will concisely review the **formulation of the problem in the case of short term and long term forecasting** and especially highlight the specific decisions this is meant to support. In technical terms, we will seek to link predictive analytics (i.e., forecasting) to the prescriptive and decision support layer.

An extensive description on all underlying issues is included in Deliverable D 7.1 and will not be reiterated here. Below, we will just concisely summarise the key decisions identified and also refer to some additional insights that have emerged in the meanwhile that deserve some consideration.

2.1 The short term use case (buildings and aggregate demand) electricity demand

Forecasting can provide valuable insight on the operation of a building and can help reveal a number of issues associated with it, which may demand our attention. For this to happen, one typically compares the forecast with the actual consumption that occurred in a given period. Should the latter be appreciably higher, this is then a clear signal of something “going wrong”.

In D 7.1 we have in detail identified that forecasting is highly important in three directions (systems, behaviour, and demand response schemes)

- For energy systems, forecasting can support **decisions related to predictive maintenance**, leading to a reduction of the building services' downtime/cost of maintenance.
- For behavioural issues, forecasting can support **decisions related to changes of operational settings and especially user behaviour change**. For example if the deviation noted is tracked down to excess heating due to high thermostat settings this may trigger a decision to lower the settings an effort towards behavioural change, i.e., trying to communicate to users the need to be cautious about this practice as it has an important impact upon consumption.
- For demand response schemes, forecasting would support the **evaluation of the various pricing schemes offered and assist in analysing the associated risks.**

Explainability is then defined, **as all approaches aim at presenting these decisions in a user friendly way to their recipients.** The figure below illustrates the three above decision support scenarios and therefore the associated explainability.

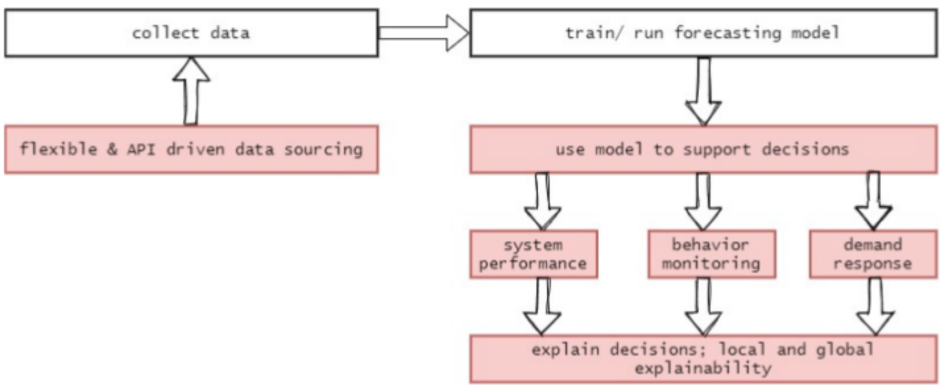


Figure 1: Three types of demand forecasting based decisions in the short term use case

2.1.1 Revisiting the short term use case in the light in newer (post D 7.1) information

In the period since the formulation of the exact requirements in D 7.1 and especially in the course of interaction with stakeholders, it was highlighted that from the three above scenarios, and from the explainability point of view

Demand response carries the greatest and most innovative potential for explainability as it is there where the inherent uncertainty prevents such schemes from taking off to their full potential. Thus, **explanations, particularly in this case could unleash significant value.**

2.2 The long term use case (country level) electricity and final energy consumption demand

In the case of country/ long term level forecasting the key pertinent problem that was reported in D 7.1 was

- to support **decisions leading to a specific CO_2 emission reduction goal**

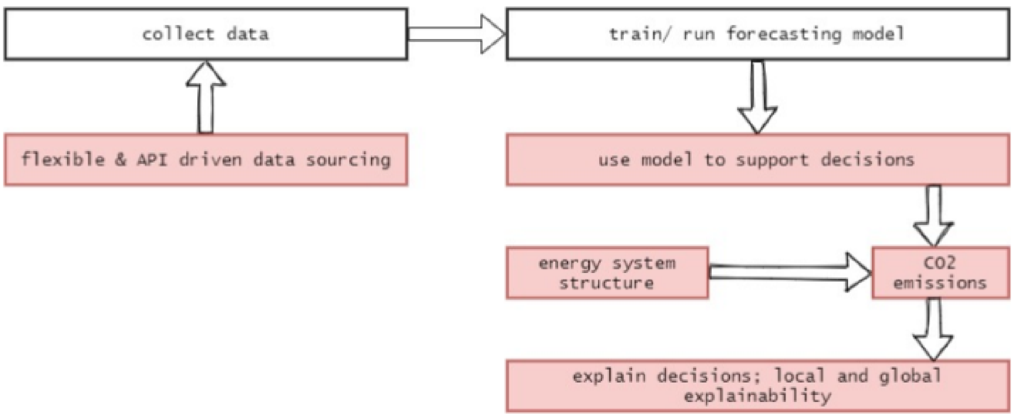


Figure 2: Country level forecasting decisions

2.2.1 Revisiting the long term use case in the light of newer (post D 7.1) information

Country level forecasting (electricity and final energy consumption) is typically addressed separately in its three key ‘domains’; **residences, transport and industry.** In our early modelling trials we have also treated it and published findings (discussed below) in this perspective.

However, during consortium discussions an interesting and value adding idea has been raised, specifically by Un. Tartu. Instead of treating each of the three domains separately to consider them all all together and formulate decisions collectively across all three domains, as a so-called **ensemble model counterfactual.** This is shown in the figure below.

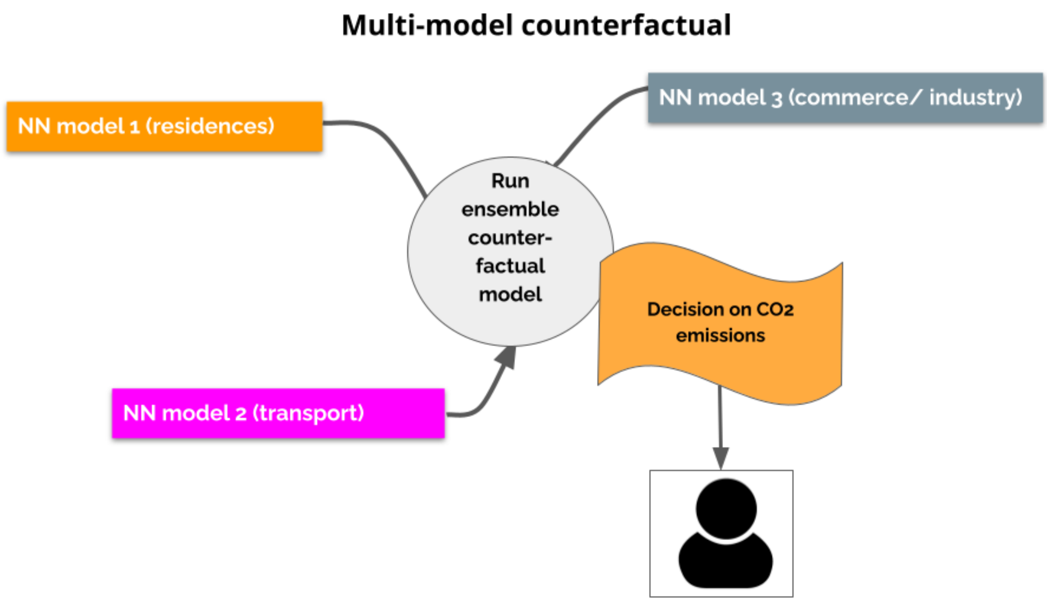


Figure 3: Ensemble model counterfactual for energy consumption and CO2 emissions

3. Explainability in the energy literature and the TRUST-AI approach

In D 7.1 we have extensively referred to the types of data used in the literature for demand forecasting in both long and short term (paragraph 3.3, page 19- 28). This discussion will not be reiterated here; When discussing below the modelling carried out we will present the TRUST-AI specific approach with some remarks drawn from the literature.

In this section we will briefly discuss a review **of explainability approaches that have appeared in the energy forecasting literature** and also present the approach as to what exactly will be done in TRUST-AI as regards explainability.

3.1 Explainability in the energy demand forecasting literature

Explainability/ Interpretability has not yet received any significant attention in the energy literature. Following an exhaustive review we can safely claim that **our early approach for interpretable county level forecasting**[1] **was the first ever to be published for forecasting in this particular long term context.**

[1] Sakkas, N.; Yfanti, S.; Daskalakis, C.; Barbu, E.; Domnich, M. **Interpretable Forecasting of Energy Demand in the Residential Sector**. Energies 2021, 14, 6568. https://doi.org/10.3390/en14206568

In the short term forecasting domain, only after 2019 did the term explainability show up in the literature, in a rather exploratory way. And only very recently, in 2022 more robust explainable approaches appeared in the literature of electricity forecasting problems.

In a first approach[2] the authors state that *"there are several attempts to explain the result of deep learning through the analysis of the* ***input attributes that influence the prediction****, but they lack appropriate explanation because of ignoring the time-series property of the input data. In this paper, we propose a* ***deep learning model to explain the impact of the input attributes*** *on the prediction by taking account of the long-term and short-term properties of the time-series forecasting."*

In a second one[3] the authors claim the following: "*our prediction model aggregates both consumption and weather information and feeds them to the embedding proposed layers in order to extract the temporal and environmental hidden features. Afterwards, we established an LSTM-based neural network model to forecast energy consumption. The energy consumption values generated by our model are evaluated and analysed by several error metrics. Finally, to increase our model's trust,* ***we rely on an agnostic method (ad-hoc and causality-based) to explain the generated predictions. However, towards scalability, the embedding layer in our system makes us lose the traceability of the original features. Thus, none of the well-known explainability frameworks (LIME, SHAPE) can be applied"***

3.1.1 Discussion

The need for explainable approaches is clearly emerging. We would argue that the full potential is not acknowledged in the literature. Explainability is in its first steps and as such, it is treated as a scientific goal, without a full realisation of the underlying business cases it may serve. For example, the fact that **explainability is a high priority and requirement for demand response practices to take off does not surface in the literature**.

The approaches pursued are tied to the dominant deep learning/ LSTM (long short term memory) modelling paradigm. The focus is on prioritising feature importance, which is indeed important, but **ideas such as genetic programming and derivation of symbolic expressions are, for the moment, completely lacking**.

TRUST AI bears the potential to provide significant input to this emerging discussion and lay out novel methods and models to approach the issue of short term forecasting. **Benchmarking performances of these approaches against the currently dominant LSTM approach would greatly add to the credibility of the results**.

---

[2] Jin-Young Kim & Sung-Bae Cho (2022): **Predicting Residential Energy Consumption by Explainable Deep Learning with Long-Term and Short-Term Latent Variable**s, Cybernetics and Systems, DOI: 10.1080/01969722.2022.2030003

[3] Mouakher, A.; Inoubli, W.; Ounoughi, C.; Ko, A. **EXPECT: EXplainable Prediction Model for Energy ConsumpTion**. Mathematics 2022, 10, 248. https://doi.org/ 10.3390/math10020248

As discussed we can also take **long term forecasting explainability one step further, by introducing the ensemble counterfactual model** (see Figure 3).

3.2 Explainability in the energy use case in TRUST-AI.

The figure below[4] provides for a categorisation of all types of explanations; [b], [c] and [d] are collectively referred to as global or model level explanations whereas [e] and [f] are the instance of local level explanations.

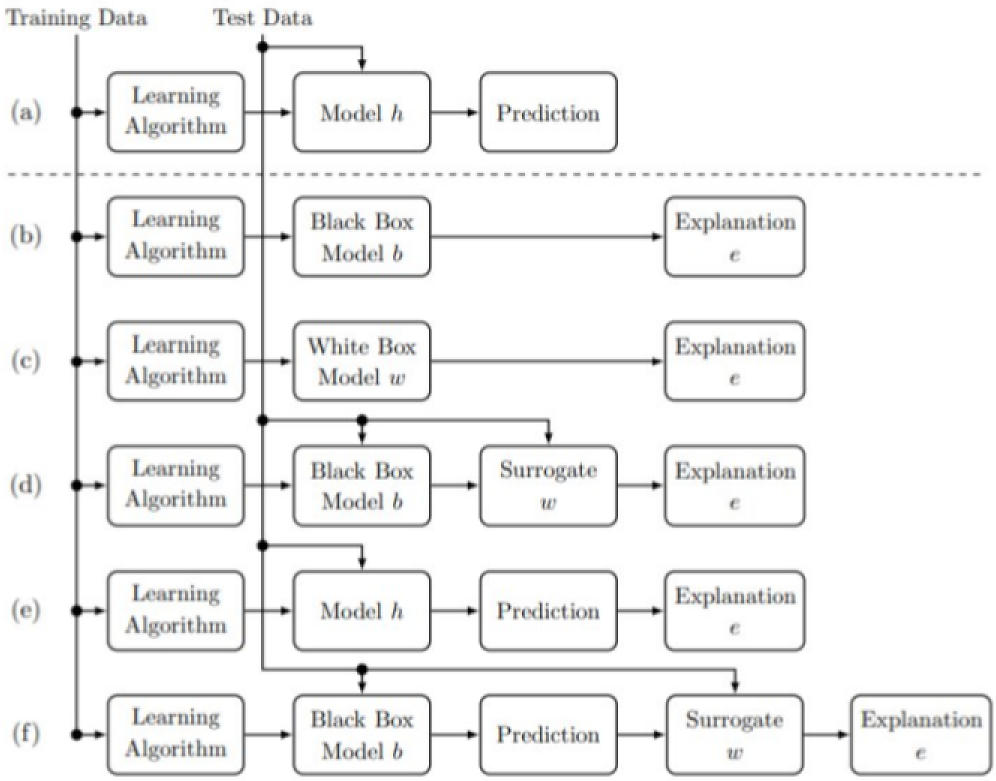


Figure 4: Literature categorisation of explainability. **Where does the value for energy use case lie? Adapted from Nadia and Marco (2021)**

What exactly is pertinent for the energy use case? We will discuss this point with reference to the illustration of Figure 4 and based on literature and stakeholder insights.

3.2.1 Global level (model) explainability [b], [c], [d]

Approach [b]. This approach implies a global level explanation which is derived from a black box model. Although this is currently the dominant approach in the literature, it is not something that will be particularly experimented upon in TRUST-AI. However, due to the emphasis it receives in the literature some benchmarking may indeed be value adding.

[4] Nadia, B.; Marco, H.F. A Survey on the Explainability of Supervised Machine Learning. J. Artif. Intell. Res. 2021, 70, 1–74.

Approach [c]. This approach is about deriving a white box model and basing the explanations upon it. A white box model could, for example, be a **symbolic expression**, something that is a key consideration in TRUST-AI. Indeed symbolic expressions could provide for a very good approach in both short term and long term forecasting. Especially, in the key short term forecasting, **seasonal symbolic expressions could help highlight the seasonal differences and provide very targeted insight on the building performance**. **This would be an important innovation of a potentially important value.** Thus approach [c] is highly relevant, especially for the short term investigation

Approach [d]. This is also potentially an area of interest. For example a black-box model may outperform an interpretable one. In this case, a surrogate (e.g., GP) can be used to explain it.

As a summary, as regards **model level explainability, approach [c] is highly pertinent, especially for the short term/ building level investigation, where crisp seasonal and building models could potentially arise**.

3.2.2 Local level (instance) explainability [e], [f]

This type of explainability, especially in the shape of counterfactuals, **emerged as a key requirement for building level approaches as well as for country/ long term approaches (see above Figures 2 & 3)**

As a summary, as regards **instance level explainability/ counterfactual analysis will be applied, preferably (but perhaps not necessarily) via approach [c] avoiding the black box altogether.**

Finally, one more important requirement that arose from the stakeholder discussions (as was that of **feature importance**, i.e., the ability to provide insights on how every feature of the model contributed to its accuracy/ explainability. In fact, this is a key purpose of explainability highlighted in the literature up to this day.

4. The short term/ building level solution approach

4.1 Overview

An office building (of approximately 1000 sq. m.) has been used for data collection. It is equipped with real time data monitoring technology provided by the POLIS-21 group. The respective real time dashboards are publicly accessible at:

https://wsn.wirelessthings.biz/v2/stef

The dashboard includes both a private (login required) and a public area (no login required). The data used is publicly available e.g., no particular authorisation is required. **To this extent all data used in the modelling below is publicly traceable on the above dashboard.**

The office building experienced a significant change of use due to the COVID pandemic in the years 2020/2021. Normal operation has only recently been restored, as of Jan. 2022 and continues so in the present. For this reason data collection, which will need to span over a year to cover all seasonal profiles, will need to keep on going till the end of the coming year. At this moment**, a six month dataset, corresponding to the winter (Jan-Mar) and spring months (Apr-Jun) has been collected and used for the modelling exercise below.** Obviously this activity will carry on till the end of 2022 **to reach conclusive evidence as regards models' benchmarking (accuracy and explainability).**

4.2 Feature engineering

We will review below the data/features extracted and used in the case of the short term building level forecasting. We will start from a concise literature review on the currently most popular approaches for the particular problem.

4.2.1 The State of the Art

In a recent review paper[5] where all current approaches in electricity forecasting are considered, the results- as far as modelling methods and features considered are concerned- are shown in the table below.

Table 4.4
Summary of ANN with swarm intelligence.

Number	Techniques	Objective	Type of inputs	References
1	ANN with PSO	To improve the forecasting precision and speed	Historical Load, Weather, Temperature and type of date	[90]
2	ANN with Chaos PSO with adaptive inertia weight	To enhanced the searching quality of forecasting	Historical Load, Meteorological data	[91]
3	BP neural network with PSO	To improve the learning speed of network and forecasting precision	Historical Load, Weather data	[92]
4	RBF neural network with PSO	To improve the precision of electric power system short term load forecasting	Historical Load, Weather condition	[93]
5	RBF neural network with ARIMA and ACO	To improve the accuracy of load forecasting	Historical Load	[94]
6	Fuzzy neural Network with PSO	To improve the accuracy of load forecasting	Historical Load	[95]
7	BP neural network with PSO	To improve the weight and threshold values	Historical Load	[96]
8	Feed Forward neural network with PSO	To forecast electrical energy consumption of equipment maintenance	Historical Load, Weather data	[18]

Table 5: The state of the art on models and features used in building electricity forecasting

[5] M. Daut et al, "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods A review", Renewable and Sustainable Energy Reviews 70 (2017) 1108–1118 http://dx.doi.org/10.1016/j.rser.2016.12.015

From the table above **as far as modelling** is concerned one can conclude, according to this paper, that:

- **Artificial Neural Network (ANN) approaches are by far the dominant** approach in the literature
- **Hybrid approaches whereby ANN is combined with particle swarm organisation** (PSO) techniques are also been increasingly used
- As laid out in the above section, there **are no explainability focussed approaches** (e.g GP/ symbolic expressions, etc.) reported in the literature.

From the table above **as far as feature engineering** is concerned one can conclude that approaches typically rely on

- **historical load data**
- **weather data** (mostly temperature & wind affecting losses and therefore *heating/ cooling* as we well as and cloud data affecting *lighting*)
- **indoor temperature**
- **type of date** (mostly the distinction between holidays and work-days)

In D 7.1, page 19 we have presented a more detailed view on features; indeed as it is shown there many other approaches have been suggested in the literature; however after two three decades of efforts in the area the above table presents some rather conclusive evidence as to what really makes sense in terms of features.

4.2.2 The opted feature approach

Here are some comments with regard to the features presented and prioritised in the above table.

- Historical data are important and have been consistently used in the modelling approaches. Also, because the trial building is equipped with smart meters the sourcing of this type of data is easy. **Therefore historical data will indeed be used in the modelling approach.** As this type of data is typically sourced every 3-4 minutes, some pre-processing will be due in order to reach hourly values of energy consumed. It should be recalled that the hourly resolution is the target one. **Pre-processing may also be required** to deal with possible data loss incidents as well as wrong and very high values that may due to some technical issue arise. This pre-processing has been included in the data service, the https://ds.leiminte.com data sharing platform that has been implemented in D 1.2 as an open data sharing platform with open pull/ push APIs.

  Issue to decide upon: Although history is a key feature in the forecasting process, the number of days one needs to look back in time is an issue to “look” back in time is included in terms of something that literature has converged upon. Typical approaches include three days, and one week. Additionally we would propose to also consider the same day of the past week.

At this moment the **weekly pattern** is used. Also, ‘day of week’ has been considered either in supplement or on its own.

- **Weather data will also be used;** although in both buildings in situ collection is in place we will prefer to use the connectivity to the local weather station to source the weather data. WT uses the weatherstack provider for such purposes. By weather data we will refer to temperature and wind. At a later and more refined stage **we may also consider cloud conditions** to the extent these may affect lighting consumption. We can then easily retrieve the historical cloud data pertaining to our study period and use them in the models. The image below illustrates the respective part of the dashboard. Humidity will not be used; only temperature and wind data

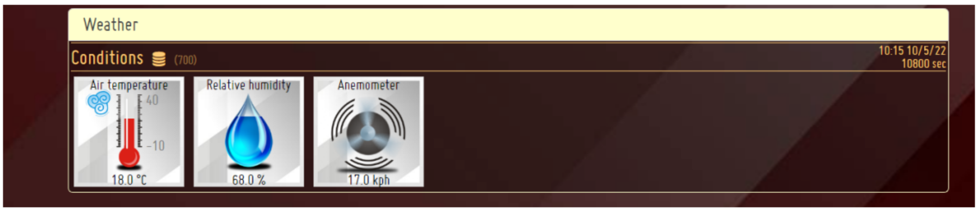


Figure 5: Weather data sourcing from weather service connected to (http://weatherstack.com)

- Lastly, **we will use indoor temperature data**, sourced from in situ sensors, which report their data, *only* when required, e.g., every time there is some significant temperature change; in this way the battery lasts over at least a year. Indeed, temperature data do not show very often in the feature set pertaining to the modelling of our scope. However indoor data has an important advantage, especially pertinent in our TRUST-AI focus. **It is the only feature that can be acted upon, and can therefore participate in counterfactual based decisions.** Due to this unique quality we will use indoor temperature.
- **As this modelling is based on winter data it seems unlikely that calendar data (day of the year and/ or the hour of the day) would make much sense.** This should be included and evaluated when a yearly pattern is available. However, in the current- winter- modelling as mentioned above, the ‘day of the week’ could be interesting to investigate. **In the final solution we look forward to integrating a calendar functionality** that will account also for holidays and will be customised for any particular site that will therefore completely streamlining the use of calendar data.
- In the current modelling we will restrict to **energy and not cost forecasting**. Cost forecasting will be important in the second stage and will require pricing data to be incorporated. This, in some cases may be trivial but in the general case it can be very complex as it may be informed by consumption levels and other information. The **final solution will have embedded a pricing tariff functionality**, custom for any particular region that will therefore completely streamline the use of such pricing data in the cost forecasting.

The following table summarises the discussion both for the early as well as for the final solution. The additional features are shown in italics

Features used for early trials	**Features used for final solution**
history consumption loads; various trials ● seven days ● same day past week	history consumption loads; various trials ● seven days ● same day past week
calendar data ● day of the week ● hour of the day	calendar data ● day of the week ● hour of the day ● *day of the year* ● *holidays*
weather data ● temperature ● wind	weather data ● temperature ● wind ● *cloud coverage*
indoor conditions ● temperature	indoor conditions ● temperature
	flexible pricing tariffs for cost forecasting

Table 1: Features for trial and final solution; short term forecasting

4.3 The modelling approach

Below we discuss the data/ models for the office building. Data were collected for Jan-Mar '22 and later for Apr-Jun '22. Some changes were introduced in the second batch of data. **All modelling was done in jupyter notebooks all of whom are publicly online at**

https://drive.google.com/drive/folders/1Y7mwyrW_4FsKnHHWMHhzmpBr7XPZ4-U1?usp=sharing

In this folder besides the notebooks also some denormalized result validation is presented to have a more hands on feeling of the forecasting performance.

Real time data from the WT were extracted from the BMS. Data was cleaned and put on the same resolution, that of an hour.

In the final solution the data will be pushed via the API in the open DS platform (ds.leiminte.com) and pulled again via the API into the final solution. Ideally all this cleaning should be done automatically by the API.

4.3.1 Introduction to the modelling approaches

A number of forecasting models have been developed and the approach is discussed below. In all cases, **a 24h day ahead electricity forecasting is the output of the models**. This is the more practical and user friendly approach.

Indeed, an alternative option would be for a user running the model today, let us say at 17.00, to return the forecasting from 18.00 today till 17.00 tomorrow. However, this approach would rely on the current hour and this could lead to potential confusion. Instead, if we run the forecasting for **tomorrow**, **always for the timeframe between 00.00- 23.00,** then the result will always be the same, regardless of the time the model is run. This is superior in terms of user friendliness while not compromising the potential use cases of the forecasting results which typically relate to the next day (and beyond) load switching or other behavioural change.

At first, a baseline/ naive model needs to be defined; such a model shall serve as a **baseline to investigate the benefits in terms of accuracy when introducing feature combinations as discussed above.**

Baseline: In the baseline model, the forecasting is just as the last day consumption. Thus, if the naive forecasting is done on some time of today (let us assume it is a Thursday [t]) it will provide forecast for tomorrow (Friday, [t+1]) and the result will simply use the yesterday consumption data (Wednesday, [t-1], from 00.00-23.00)

Additionally it is important to note that these models have been developed via two distinct approaches as follows:The first one will be based on LSTM (long short term memory) approaches which is currently among the most frequently used AI approaches in the type of problems under consideration. **The second one will be based on GP (genetic programming) approaches that are central in the TRUST-AI approach and that may potentially offer a much more transparent and explainable model.** The extraction of symbolic expressions, pertaining to the particular building is a further added value potentially resulting from the GP approach. Such expressions can be pivotal for building benchmarking, and provide new outlooks on building characterisation.

Below is a description of the various models that have been elaborated.

4.3.2 Models based on the data from the <u>winter</u> months (Jan- Mar 2022)

Models developed for this period are tabulated as shown below: Metrics for best performing models (shown in yellow) will be discussed in the downstream.

CODE	WINTER	SPRING	FEATURES	MODEL TYPE	BEST PERFORMING
1	YES	NO	weekly consumption history [t-1, t-8]	LSTM	
2a	YES	NO	weekly consumption history [t-1, t-8] day of week [t+1]	LSTM	
2b	YES	NO	weekly consumption history [t-1, t-8] indoor [t+1]	LSTM	

2c	YES	YES	weekly consumption history [t-1, t-8] day of week [t+1] indoor [t+1]	GP/ LSTM	
2c'	YES	YES	weekly consumption history [t-1, t-8] indoor [t+1] day of week [t+1] hour of day [t+1]	GP/ LSTM	LSTM WINTER LSTM SPRING
3a	YES	NO	weekly consumption history [t-1, t-8] day of week [t+1] outdoor temp [t+1] wind [t+1]	LSTM	
3a	NO	YES	weekly consumption history [t-1, t-8] day of week [t+1] outdoor temp [t+1]	LSTM	
3a	YES	YES	weekly consumption history [t-1, t-8] day of week [t+1] outdoor temp [t+1] hour of day [t+1]	GP	
3a'	YES	YES	weekly consumption history [t-1, t-8] ABS (indoor- outdoor) [t+1] day of week [t+1] hour of day [t+1]	GP	GP WINTER
3a'	NO	YES	weekly consumption history [t-1, t-8] outdoor [t+1] day of week [t+1] hour of day [t+1]	LSTM	
3a''	YES	YES	weekly consumption history [t-1, t-8] ABS (indoor- outdoor) [t+1] day of week [t+1]	GP	GP SPRING
3b	YES	NO	weekly consumption history [t-1, t-8] day of week [t+1] outdoor temp [t+1]	LSTM	
3c	YES	NO	weekly consumption history [t-1, t-8] day of week [t+1] wind [t+1]	LSTM	
4a	YES	YES	weekly consumption history [t-1, t-8] day of week [t+1] outdoor temp [t+1] indoor [t+1]	LSTM	
4b	YES	NO	weekly consumption history [t-1, t-8] day of week [t+1] wind [t+1] indoor [t+1]	LSTM	

4c	YES	NO	weekly consumption history [t-1, t-8] day of week [t+1] outdoor temp [t+1] wind [t+1] indoor [t+1]	LSTM	
4a'	YES	YES	weekly consumption history [t-1, t-8] day of week [t+1] outdoor temp [t+1] indoor [t+1] hour of day [t+1]	LSTM	
5	YES	NO	same day of past week consumption [t-6]	LSTM	
6a	YES	NO	same day of past week consumption [t-6] outdoor temp [t+1]	LSTM	
6b	YES	YES	same day of past week consumption [t-6] outdoor temp [t+1] day of week[t+1]	LSTM	
7a	YES	NO	same day of past week consumption [t-6] outdoor temp [t+1] indoor [t+1]	LSTM	
7b	YES	YES	same day of past week consumption [t-6] outdoor temp [t+1] indoor [t+1] day of week [t+1]	LSTM	
7c	YES	NO	same day of past week consumption [t-6] indoor [t+1]	LSTM	
7d	YES	NO	same day of past week consumption [t-6] indoor [t+1] day of week [t+1]	LSTM	
8	YES	YES	indoor [t+1] day of week [t+1]	STS	STS WINTER STS SPRING
8'	YES	YES	indoor [t+1] day of week [t+1] hour of the day [t+1]	STS	

4.3.3 Discussion on models based on the data from spring months (Apr- Jun 2022)

Following evaluation of the first three months analyses (winter data) some decisions were made on the following grounds:

First, we needed to stop investigating models that did not seem feature-wise to perform well so that we would be able to focus on less and more promising models. To this extent in the spring period we restricted and developed only the following models.

Detailed LSTM/ STS model list for the spring data (first part)

1. Model 2c : trained on: weekly consumption history [t-1, t-8], day of week [t+1], indoor
2. Model 3a: trained on: weekly consumption history [t-1, t-8], day of week [t+1], outdoor [t+1]
3. Model 4a: trained on: weekly consumption history [t-1, t-8], day of week [t+1], outdoor [t+1], indoor [t+1]
4. Model 6b: trained on: same day of past week consumption [t-6], outdoor [t+1], day of week[t+1]
5. Model 7b: trained on: same day of past week consumption [t-6], outdoor [t+1], indoor [t+1], day of week [t+1]
6. Model 8: STS model

Second, we considered it useful to start including the impact of 'hour of day'. To this extent the following new models were developed. These models are coded as above and we use the (') to denote the additional feature used (hour of day)

Detailed LSTM/ STS model list for the spring data (second part)

1. Model 2c': trained on: weekly consumption history [t-1, t-8], indoor [t+1], day of week [t+1], hour of day [t+1]
2. Model 3a': trained on: weekly consumption history [t-1, t-8], outdoor [t+1], day of week [t+1], hour of day [t+1]
3. Model 4a': trained on: weekly consumption history [t-1, t-8], indoor [t+1], outdoor [t+1], day of week [t+1], hour of day [t+1]
4. Model 8': STS model (features used), hour of day [t+1]

Third, we needed to look further into GP models, both for winter and spring. The plan was to develop GP models along the following:

- Consider model 2c
- Consider model 2c with the spring data and also introduce the hour of the day (model 2c')
- Consider model 3a' (like 3a but with the spring data and also introducing the hour of the day)
- Consider two additional models, variants of above 3a' (3a'' and 3a''') where **both indoor and outdoor temperatures would be included via a single feature, their ABS difference**, which is essentially what drives the HVAC consumption.

4.3.4 Data and Model public Access (building case)

Data and Models are and will remain publicly available and update continuously at

https://drive.google.com/drive/folders/1Y7mwyrW_4FsKnHHWMHhzmpBr7XPZ4-U1?usp=sharing

There are several folders therein. Models are included in the folders

- GP
- NN LSTM
- STS

Whereas, data is included in the folder

- DATA

4.3.5 Seasonal Analysis

A seasonal analysis of the two trimester datasets is presented below. This allows us to get a better feeling of the data and how they evolve in time.

a. Plotting Winter/ Spring data

The below plot shows active_electricity, indoor temperature and outdoor_temperature for winter and spring data, respectively

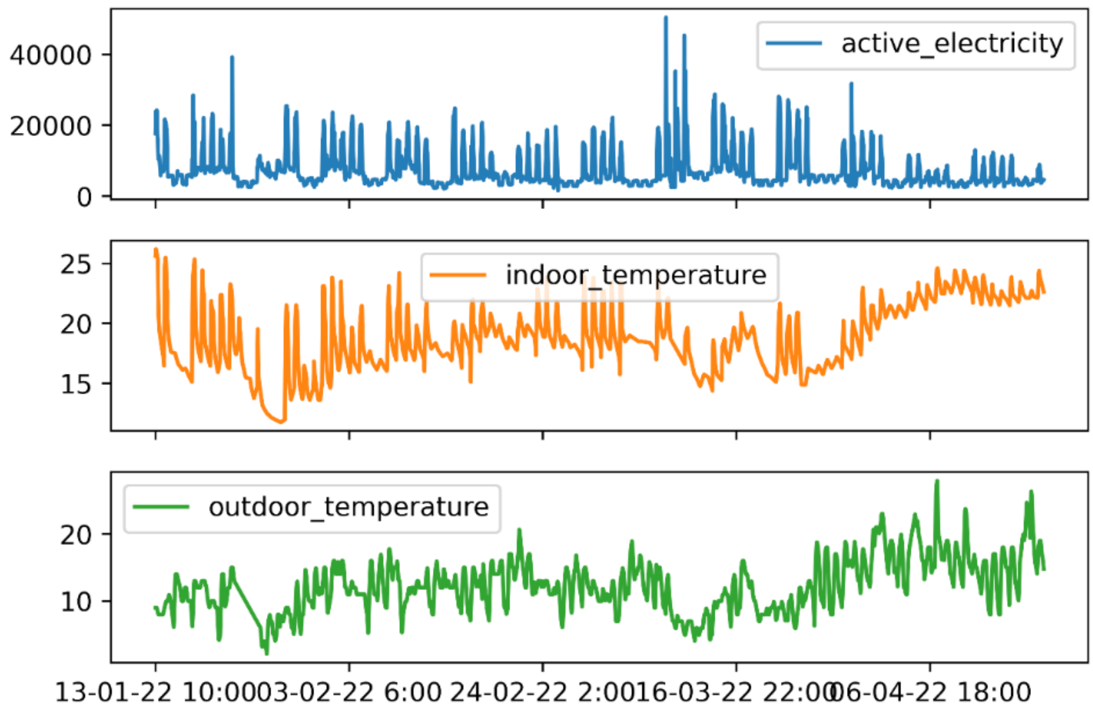


Figure 6: Plotting the winter data

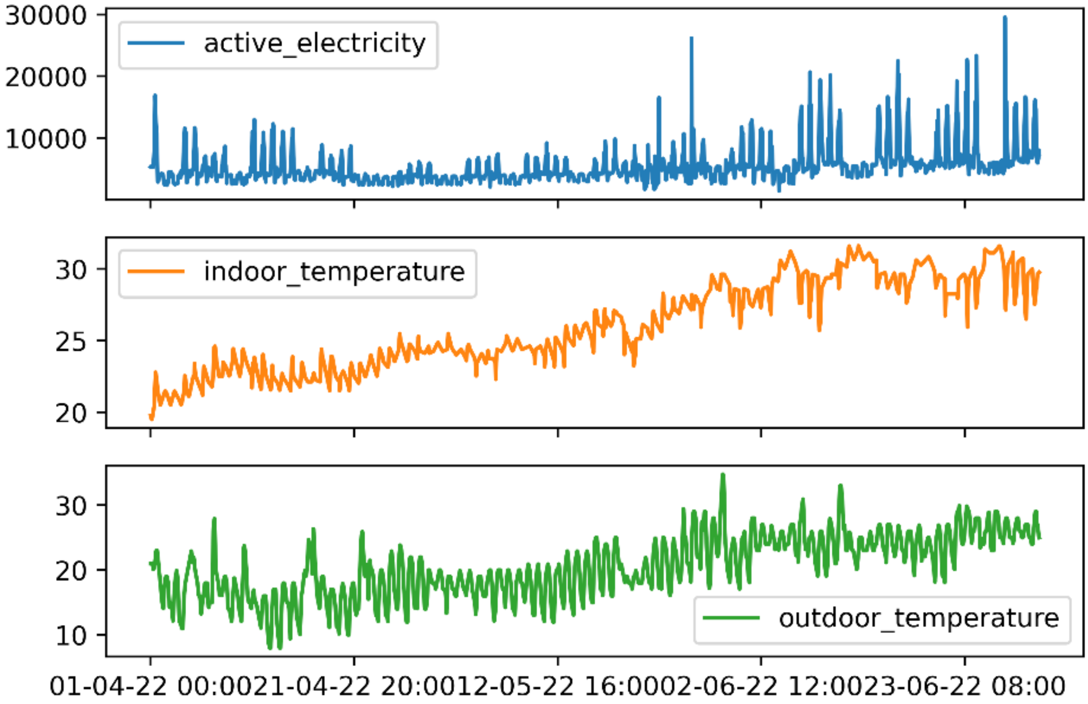


Figure 7: Plotting the spring data

b. Daily patterns (seasonality)

The daily seasonality patterns are shown below for winter and spring data respectively. As one can see the winter data has a **very strong repeating seasonal pattern**, which almost repeats throughout. Only at the onset of April the pattern has shorter peaks.

In spring the seasonal pattern between the end of April to mid May shows lower consumptions, the reason being clearly visible in Figure 9(as indoor temperature is at a level that no significant cooling/heating is needed). From the end of May the spring seasonal patterns in electricity consumption becomes somewhat similar to winter data as the use of cooling appliances increases with increase in indoor temperature, just like heating appliances caused higher consumption in winter.

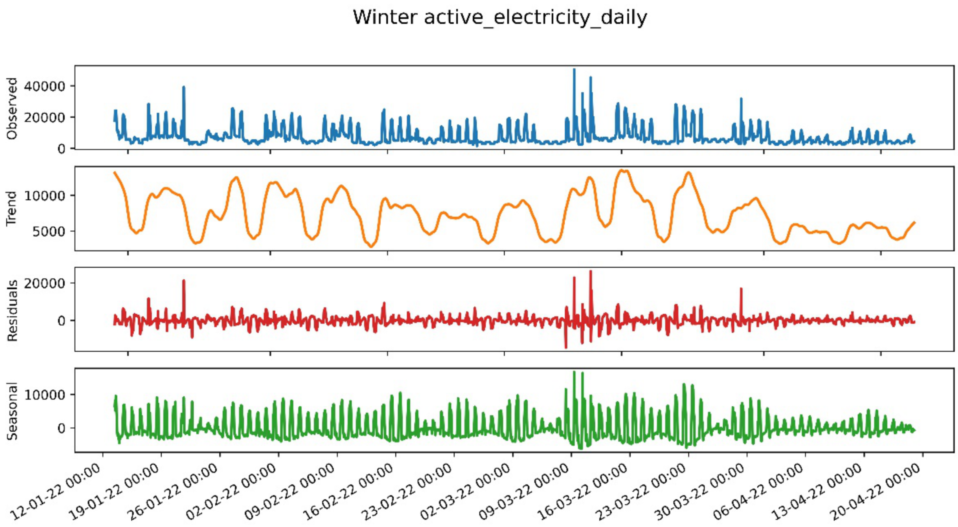


Figure 8: Winter data active electricity data decomposition for viewing daily seasonality patterns

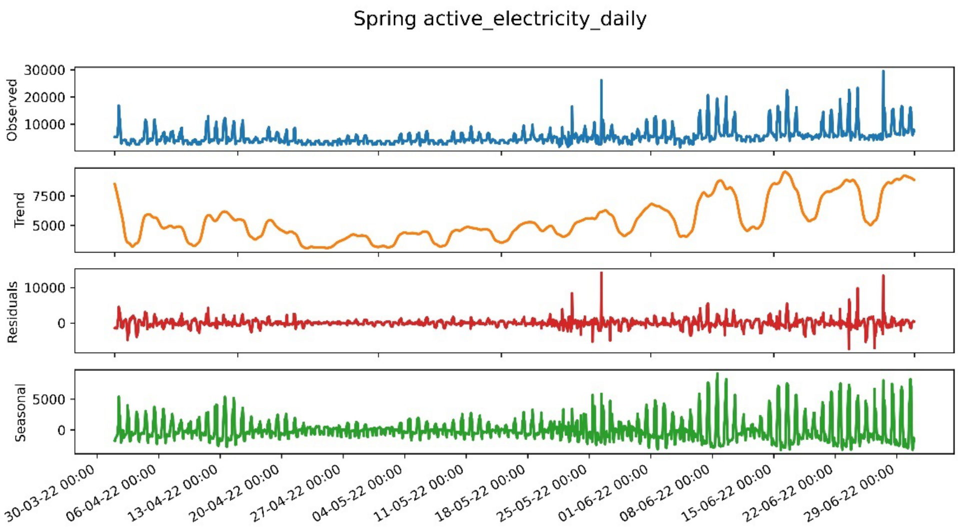


Figure 9: Spring data active electricity data decomposition for viewing daily seasonality patterns

c. *Weekly patterns (seasonality)*

The weekly seasonality patterns are shown in figures 11 and 12, for winter and spring data respectively. As we see here the winter data consumption with weekly periods also exhibits a very strong repeating seasonal pattern, which almost repeats throughout. April has differences in the pattern, as also observed in daily patterns. Here too, in spring consumption data between April end to mid May shows lower consumptions like in daily seasonality. From the end of May the seasonal patterns in spring electricity consumption becomes almost similar to winter data. The trend component shows that in winter the consumption decreases as spring approaches, whereas in spring the trend has an increasing pattern, because the indoor temperature also exhibits a similar behaviour leading to increased electricity consumption by cooling appliances.

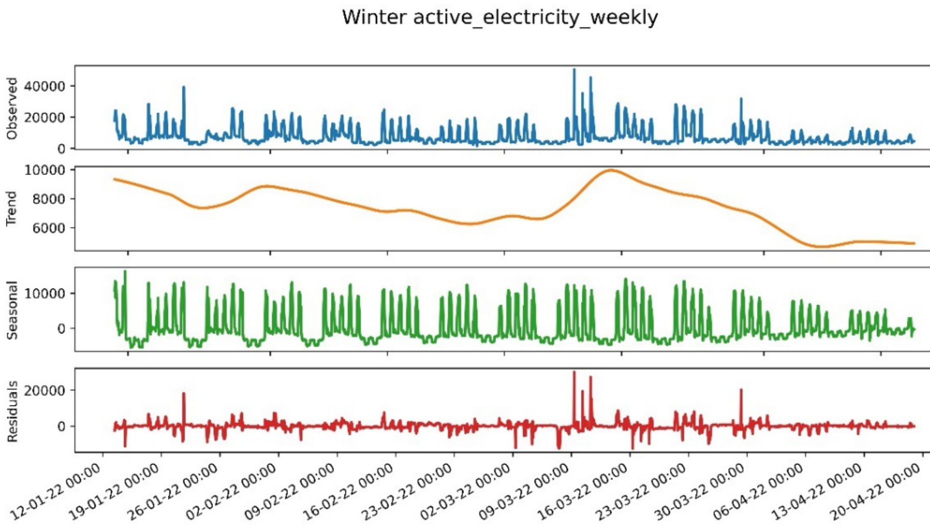


Figure 10: Winter data active electricity data decomposition for viewing weekly seasonality patterns

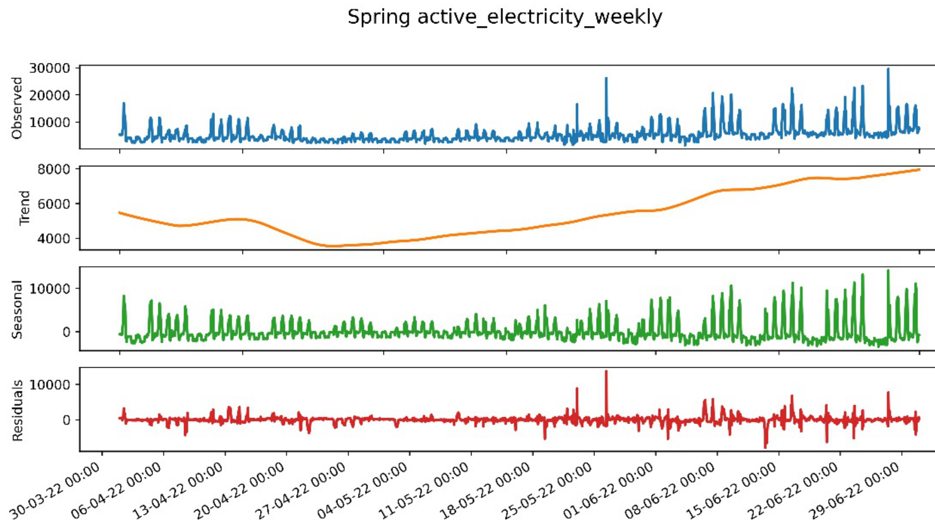


Figure 11: Spring data active electricity data decomposition for viewing weekly seasonality patterns

d. Decomposition over week / day

As seen in Figure 12, the seasonal component plot here shows that the electricity consumption exhibits an expected seasonal pattern showing valleys on weekends (2nd and 3rd April) due to lower consumption on weekends in offices.

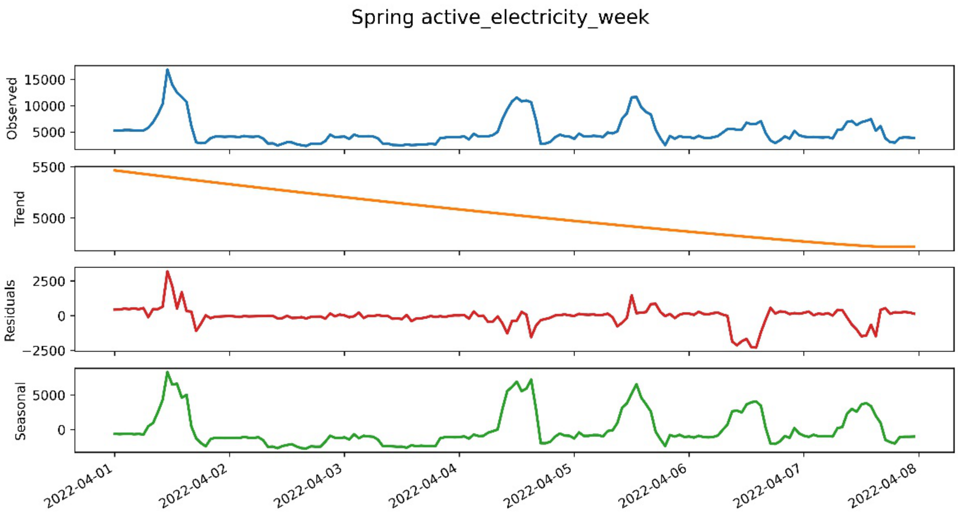


Figure 12: Spring electricity consumption decomposition over one week (01-08 April)

As seen in Figure 13, the seasonal component shows increased electricity consumption during peak office hours and valleys during non-office hours.

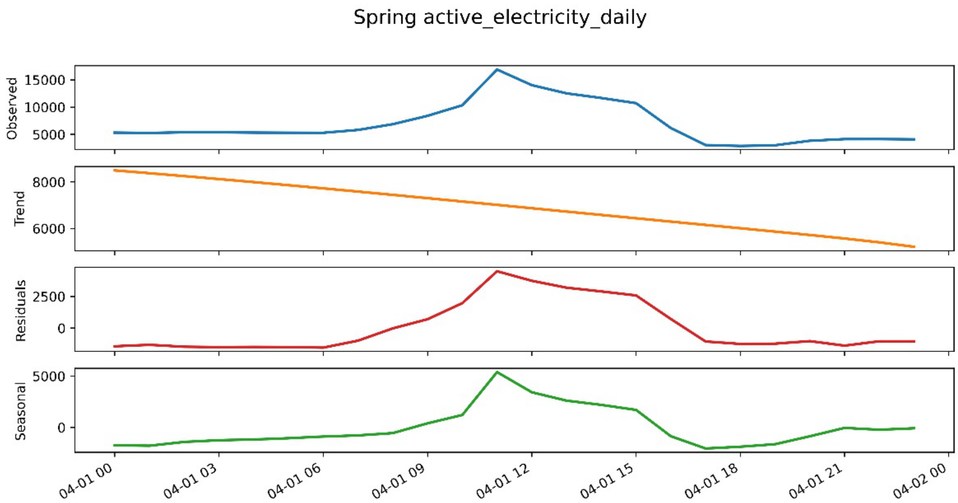


Figure 13: Spring electricity consumption decomposition over a day (1st April)

4.3.6 The LSTM/ STS baseline

We illustrate below results extracted from the 2c notebook. This is one of the preferred approaches as it includes an actionable feature **'indoor'**. illustrated, it includes various sub models, for example alone a baseline, naïve model, a linear model, a LSTM model and several more.

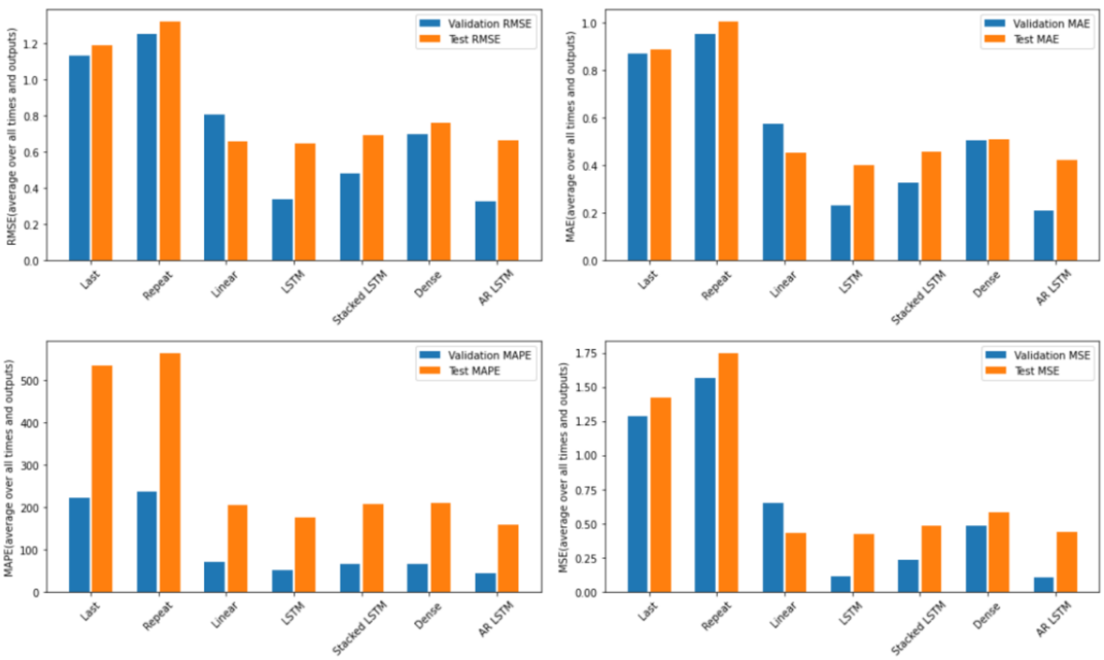


Figure 14: LSTM 2c model

Overall it can be seen that there is a significant increase in prediction accuracy when compared with the naive models.

4.3.7 Benchmarking the LSTM, the GP and STS models

A number of metrics (RMSE, MAE, MAPE, MSE) has been used for the evaluation of all models. Here below we will present some key results and graphs of these models, mostly focussing on the accuracies achieved in each case. A discussion closes the section.

We present below some comparative results as regards the 2c model as for three different methods

- The NN/ LSTM based approach
- A GP based approach
- A STS (structural time series) based approach

```
In [ ]: print('Metrics LSTM for only active_electricity denormalized')
        print('MAPE:',tf.keras.backend.mean(tf.keras.metrics.mean_absolute_percentage_error((inputs[:,:,0]*train_std[0])+train_mean[0],(
        print('MAE:',tf.keras.backend.mean(tf.keras.metrics.mean_absolute_error((inputs[:,:,0]*train_std[0])+train_mean[0],(predictions[
        print('MSE:',tf.keras.backend.mean(tf.keras.metrics.mean_squared_error((inputs[:,:,0]*train_std[0])+train_mean[0],(predictions[:,
        print('RMSE:',math.sqrt(tf.keras.backend.mean(tf.keras.metrics.mean_squared_error((inputs[:,:,0]*train_std[0])+train_mean[0],(pre

        Metrics LSTM for only active_electricity denormalized
        MAPE: 47.741493
        MAE: 2746.022
        MSE: 18165088.0
        RMSE: 4262.0520879031965
```

Figure 15: LSTM accuracies/ 2c model

```
In [55]: #denormalize gp results using inverse scaler
denormalised_results_gp=compute_metrics(y_weekly_scaler.inverse_transform( gp_weekly_exp.observed_vals), y_weekly_scaler.inverse_
print('Denormalised Metrics for GP model: ')
print('MSE: ',denormalised_results_gp[0])
print('RMSE: ',denormalised_results_gp[1]),
print('MAE: ',denormalised_results_gp[2])
print('MAPE%: ',denormalised_results_gp[3])

Denormalised Metrics for GP model:
MSE:  24755553.76
RMSE:  4975.5
MAE:  2695.45
MAPE%:  40.41
```

Figure 16: GP/ 2c model

```
In [25]: print('Metrics sts model over test data')
print('MAPE:',tf.keras.backend.mean(tf.keras.metrics.mean_absolute_percentage_error(demand.to_list()[len(demand_forecast_mean):],
print('MAE:',tf.keras.backend.mean(tf.keras.metrics.mean_absolute_error(demand.to_list()[len(demand_forecast_mean):],demand_fore
print('MSE:',tf.keras.backend.mean(tf.keras.metrics.mean_squared_error(demand.to_list()[len(demand_forecast_mean):],demand_foreca
print('RMSE:',math.sqrt(tf.keras.backend.mean(tf.keras.metrics.mean_squared_error(demand.to_list()[len(demand_forecast_mean):],de

Metrics sts model over test data
MAPE: 63.12593051974821
MAE: 3401.4603132419093
MSE: 23813636.4015746
RMSE: 4879.9217618292405
```

Figure 17: STS accuracies of a LSTM model trained on spring data

We compare below the results obtained from LSTM, GP and STS models trained and tested over spring and winter electricity consumption data, by using the feature combination that in each case appeared to be the best performing, with relatively best metrics based on the results obtained. This optimum feature set is as follows:

- for the **LSTM models**:
  - o **LSTM_Spring** (over spring data): The feature set denoted as 2c'. The LSTM models have been trained with split as per: 0-70% training set, 70-85% validation set, 85-100% test set.
  - o **LSTM_Winter**(over winter data): The feature set denoted as 2c'. The LSTM models have been trained over winter data with split as per: 0-40% training set, 40-50% validation set, 50-100% test set.

  We use a window of past 7 days consumption to predict 24 values of the next day

- for the **GP models**:
  - o **GP_Spring** (over spring data): The feature set denoted as 3a''. The GP models have been trained with split as per: 0-85% training data and 85-100% testing data.
  - o **GP_Winter** (over winter data): The feature set denoted as 3a'. The GP models have been trained over winter data with split as per: 0-50% training data and 50-100% test data.

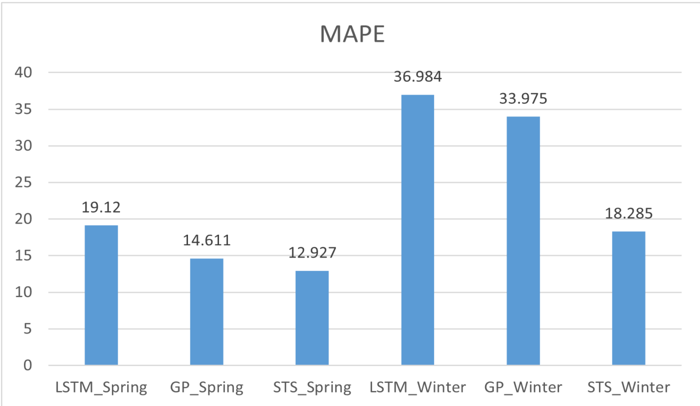
  The criteria used for fitness is RMSE.

- for the **STS** models **STS_Spring** and **STS_Winter**: the feature set denoted as 8 for both spring and winter data. The STS model's performance with a one-step predictive model is better as compared to forecasting ahead a given number of time steps.

Follows the benchmarking with regard to the MAPE, RMSE, and MAE metrics.

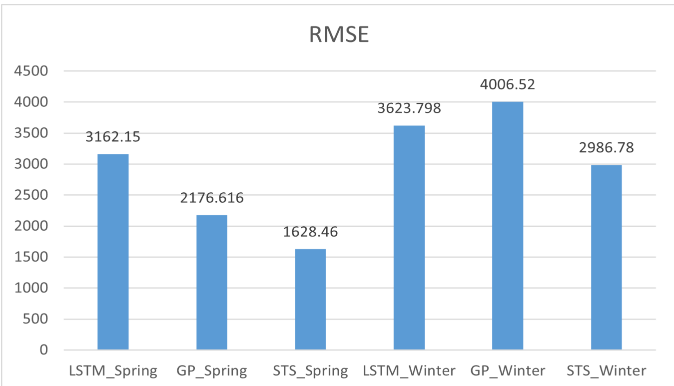
MAPE

As we can observe MAPE wise here STS one step predictive model performs best amongst the three with 12.92% MAPE for STS_Spring model and 18.25% MAPE for STS_Winter model, followed by GP, whereas LSTM shows the poorest performance as per MAPE.



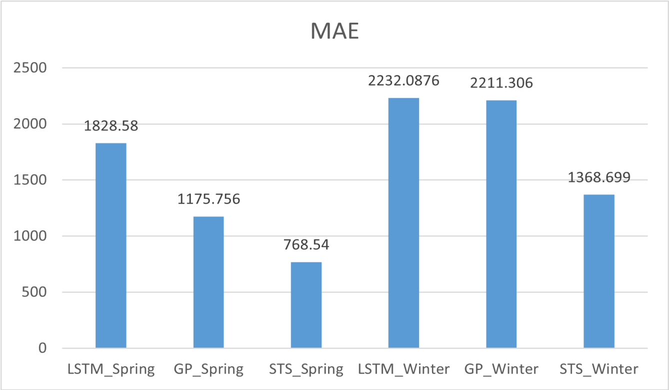
RMSE

With regard to this metric the STS_Winter and STS_Spring models again show the best performance closely followed by the GP_Spring model, the GP_Spring model lags behind STS_Spring model by 3.6%. Whereas here the LSTM_Winter model shows comparatively better performance as compared to GP_Winter model.



MAE

According to MAE, the STS_Spring and STS_Winter models again show the best performance, whereas the GP_Spring model follows and LSTM_Spring is again the model with the worst MAE among the three. Although LSTM_Winter is lagging behind GP_Winter by minimal difference as seen here.



4.3.8 Discussion and Conclusions

Following this modelling benchmarking across LSTM/GP and STS approaches as in a synopsis provided in the above figures one can **reach the following conclusions:**

- **The GP approach significantly outperforms the LSTM approach** that it typically used for the forecasting purpose. All three metrics used clearly show in this direction.
- **STS models appear to perform slightly better than the GP approach when forecasting an hour ahead. However when forecasting the full next day again GP outperforms STS.**
- **As we did not find any performance advantage on the side of LSTM so this modelling approach will be discontinued** from further experimentation. Performance is lagging well behind and if we add to this the obscurity of the approach and its black box nature there appears little sense in persisting in this direction. We may only test the so-called *transformer models*, which are considered successors to LSTM models and have a very good performance history with NLP tasks which are also textual sequences, the attention mechanism might prove helpful to model our time series data too. On the contrary, **STS will be further kept in the benchmarking as its performance is well acceptable.**
- The **symbolic expressions associated with the GP models are very long and counter intuitive**. To this extent, the explainability potential is not really obvious nor harnessed. One needs to 'prune' this expression aiming at arriving at more intuitive and workable expressions. This important exercise will be carried out when the TRUST AI framework is available, towards the end of 2022. Ideally we would be able to arrive at compact, seasonal expressions that would not significantly compromise accuracy. This however remains to be seen.
- The **whole exercise needs to be repeated in other seasons in order to reach safe conclusions applicable throughout the year**. Not all models presented above need to be re-elaborated in the coming seasons. However some new features (e.g., hour of the day) are worth exploring.

- Some **causality elaborations are currently ongoing in a parallel thread** and their results may again impact upon the select optimum feature set and the related modelling

Insights gained for next modelling steps

1. For the GP models we currently use only (+, -, *, /) operators. **It would be important to also experiment with operators** like sin, cos, min, max, square root, log or other custom ones, for getting more variations with symbolic expressions.

2. **It is suggested that this GP operator selection and elaboration should be carried out within the upcoming TRUST AI framework**, where we will be able to gain insight in the symbolic expressions, especially in the direction of pruning them into explanatory expressions. The trials could be carried on both on available winter and summer data as well as upcoming summer data (available end of September '22)

3. **Transformer models will be tested on the autumn data to see if they provide any noticeable enhancement** over LSTM which, as said above, will be abandoned in the upcoming modelling

4. For the STS Model we currently observe the performance over a one-step predictive model which looks at all observations up to time T-1 to predict time T, which is different from our window in GP and LSTM which looks at 7 days consumption history to predict. It should be changed in accordance with the GP model.

5. There are some differences in the modelling which in the next seasons need to be addressed so a fully common baseline is established. Follow some details on this aspect.

The current approach:

As per the approach used with GP models currently, we are generating one step ahead forecast but in a rolling manner, so to predict 15.04 00:00 we are using consumption and indoor temperature values from 08.04 00:00, as seen in the image of the rolling window used below. This is from the model you had shared earlier. Note: active_electricity is the target variable and active_electricity_weekly is the input variable.

```
weekly_consumption[0:5]
```

DateTime	active_electricity	indoor_temperature	outdoor_temperature	weekday	active_electricity_weekly	indoor_temperature_weekly	outdoor_temperature_weekly	weekday_weekly
2022-04-08 00:00:00	3893.025	22.514596	16.833333	Friday	5319.8625	19.777834	21.0	Friday
2022-04-08 01:00:00	3939.425	22.500000	16.166667	Friday	5271.4000	19.655545	21.0	Friday
2022-04-08 02:00:00	3889.050	22.500000	16.250000	Friday	5366.5000	19.533256	21.0	Friday
2022-04-08 03:00:00	3912.175	22.500000	16.583333	Friday	5366.5000	19.500000	21.0	Friday
2022-04-08 04:00:00	3767.525	22.500000	16.916667	Friday	5323.8375	19.624472	21.0	Friday

```
weekly_consumption[168:173]
```

DateTime	active_electricity	indoor_temperature	outdoor_temperature	weekday	active_electricity_weekly	indoor_temperature_weekly	outdoor_temperature_weekly	weekday_weekly
2022-04-15 00:00:00	3821.908333	21.979567	15.000000	Friday	3893.025	22.514596	16.833333	Friday
2022-04-15 01:00:00	3852.925000	21.935775	15.000000	Friday	3939.425	22.500000	16.166667	Friday
2022-04-15 02:00:00	3718.350000	21.891982	14.750000	Friday	3889.050	22.500000	16.250000	Friday
2022-04-15 03:00:00	3756.862500	21.830897	14.416667	Friday	3912.175	22.500000	16.583333	Friday
2022-04-15 04:00:00	3800.175000	21.763298	14.083333	Friday	3767.525	22.500000	16.916667	Friday

Current approach; weekly consumption history models

As seen in the image below, to predict 17.04 00:00 we are using consumption history from 09.04 00:00 to 16.04 00:00 (7 days consumption history). Note: active_electricity is target variable and active_electricity_192 is input variable.

```
input[input.columns[::-1]].head()
```

DateTime	active_electricity	outdoor_temperature	Day sin	Day cos	Hour sin	Hour cos	active_electricity_24	active_electricity_25	active_electricity_26	active_electricity_27	...	active_electricity_183	active_el
09-04-22 0:00	3856.9250	16.833333	-0.433884	-0.900969	0.000000	1.000000	3893.025	3884.717553	3949.582447	3981.475000	...	8403.92500	
09-04-22 1:00	3978.2625	16.166667	-0.433884	-0.900969	0.258819	0.965926	3939.425	3893.025000	3884.717553	3949.582447	...	10343.32181	
09-04-22 2:00	3978.9125	16.250000	-0.433884	-0.900969	0.500000	0.866025	3889.050	3939.425000	3893.025000	3884.717553	...	16902.16569	
09-04-22 3:00	3919.9500	16.583333	-0.433884	-0.900969	0.707107	0.707107	3912.175	3889.050000	3939.425000	3893.025000	...	14037.13750	
09-04-22 4:00	4015.8000	16.916667	-0.433884	-0.900969	0.866025	0.500000	3767.525	3912.175000	3889.050000	3939.425000	...	12530.95000	

5 rows × 175 columns

```
input[192:195]
```

DateTime	active_electricity_192	active_electricity_191	active_electricity_190	active_electricity_189	active_electricity_188	active_electricity_187	active_electricity_186	active_electricity_185	active_electricity_
17-04-22 0:00	3856.9250	3978.2625	3978.9125	3919.95	4015.80	3878.25	4021.7	3478.5	2533
17-04-22 1:00	3978.2625	3978.9125	3919.9500	4015.80	3878.25	4021.70	3478.5	2533.0	2694
17-04-22 2:00	3978.9125	3919.9500	4015.8000	3878.25	4021.70	3478.50	2533.0	2694.4	2791

So as per our current approach **we are getting the forecast value for one step, not 24 steps as a vector, which is what we are getting with LSTMs**. Gplearn does not support multi-output in an inbuilt manner, whereas neural network support vector outputs too, so multi-output support is inbuilt there.

Likely solutions:

If we want to predict 24 values multi-step output with gplearn we have two options either we use a recursive forecast strategy or a direct forecast strategy. In a recursive forecast the model will predict one time step and pass the prediction back as input in the rolling input window. One main problem here can be that errors propagate. The second option can be direct forecasting where we have 24 different models that are trained to predict 24 different timesteps given the same input window, ex: one to predict timestep 00:00, another one to predict timestep 01:00….23:00. **We will opt for the recursive approach, to that we end up with just one model on whose expressions we can focus our attention.**

5. The long term/country level solution approach

5.1 Overview

Investigations in this context aim at **policy support, in a more mid-term horizon, addressing and supporting decisions in investment planning and energy pricing**. In our approach **we have started by developing a forecasting baseline, by means of a mixture of statistics and NN modelling. Also, by introducing some first local explainability/ counterfactual approaches.**

Global GP approaches (i.e., symbolic expressions) are to be introduced in the near future.

Although in the respective literature there are a number of AI approaches pursued, it is interesting to note that the concept of interpretability and explainability has been up to this moment completely lacking. **Our work in this is introduced and published for the first time ever such notions in the scientific literature and in particular for the residential domain,** although the same concepts are currently also transformed for transport related energy.

Various possible approaches to model the explanatory variables used have been considered; the end goal extended beyond model accuracy; it expanded also to interpretability and counterfactual concepts and analysis, aiming at the development of a modelling approach that can provide decision support for strategies aimed at influencing energy demand.

Thus, the **explainability approach pursued was for the moment that of local/ instance level and more specifically about counterfactuals**. Indeed, we addressed an important, and strongly controversial literature; how can pricing affect final total consumption of electricity consumption). Price was identified as the only essentially actionable parameter of these models; therefore it is susceptible for counterfactual analysis: *what is the minimum residential energy price change required to reach a specific consumption level?*

After models have been developed also for the transport and the industry case (currently working on them), **state of the art ensemble counterfactual analysis** will be possible that will allow to address the question: *what is the minimum energy price change and in which specific sector (residences, transport, industry) required to reach a specific consumption level?* We have explained this approach in Figure 3.

Mid-term forecasting of final energy and electricity for the residential sector has been addressed in six EU countries (Germany, the Netherlands, Sweden, Spain, Portugal and Greece) based on data sourcing that is described below.

5.2 Feature engineering

Feature engineering abided by the following criteria.

- We sought to **include all major 'causes' of consumption**; we also tried to include only one feature for every 'cause', avoiding double counting. Similarly, we considered 'causes' as independently as possible, avoiding semantic overlapping as much as possible. The statistical pre-processing of the data before moving into AI was helpful in this direction.
- We **restricted the investigation to those features that are particularly relevant to our use case and its mid-term timeframe**. For example, we will not include any dematerialisation features, such as that attempted by Sun J.W. [5], on the grounds that this will not significantly manifest over the mid-term, which is our key concern here.
- We **placed a special interest in actionable features**, i.e., features we can tweak and act upon. This is important in the case of decision support. For example, weather parameters are not actionable. They may significantly affect consumption, and may therefore fulfil above criterion 1 and deserve inclusion in our models; however, from the decision support perspective they cannot be acted upon.
- Finally, **data availability was also important and posed some important constraints**. In a short-term investigation, we always have the option to generate the data we consider essential; in our mid-term timeframe, there is little chance to do so. One has to rely on data that can already be registered and trusted.

The feature selection process has been extensively presented in the paper referenced above. We will not repeat it below and restrict to just presenting the final selection without discussing the selection process.

	Features				
Type of Forecasting	**Weather**	**Size of Use**	**Intensity of Use**	**Price**	**Efficiency of Use**
Feature can be acted upon in the mid term	NO	NO	NO	YES	NO
1. Final consumption in households	HCDD	population floor area of dwellings	GDP/capita (PPP adjusted) average private household consumption floor area of dwellings per capita	Gas price with and without taxes	ODEX technical efficiency indicator
2. Electricity consumption in households	HCDD	population floor area of dwellings	GDP/capita (PPP adjusted) average private household consumption floor area of dwellings per capita	Electricity price with and without taxes	ODEX technical efficiency indicator

Table 2: Candidate features for the forecasting

Data was collected from two sources; EuroStat and the Odyssey Mure, a EU database on energy. **It is interesting to note that this extensive data collection resulted in several data sets that were publicly shared on OSF (www.osf.io).**

5.3 Modelling approaches

Running a model on so many features does not seem a prudent approach. It is certain that several of these features are strongly interrelated and therefore collinearity will be present.

The analysis that follows below aims at isolating the most promising features. Arguably no matter how much we have tried to define independent features, there will still be many interdependencies among them, something that will manifest in terms of high collinearity indices. Thus, a number of trials were necessary to end up with the minimal set of features, providing a satisfactory model while suppressing collinearity. The data approach will be discussed below.

The statistical analysis was carried out in JASP. Various feature combinations were tested and a minimal set was retained. No counter-intuitive coefficients showed up in the regression equation; no special action or elaboration was therefore necessary as regards this important aspect of the forecasting. Additionally, a further important consideration was to retain only features that entered the equations with a low p-value, signalling a good statistical significance of it. Furthermore, features that, upon inclusion, displayed a high collinearity as reported by the SVI indicator were excluded. Collinearity means that the introduction of a new feature does not introduce independent information; it is somehow already correlated with one of the other predictors.

Following this approach we achieved to reduce the candidate features to, in most cases, three. Table 5 and Table 6 in the footnote 1 reference (pages 9 and 10) paper summarises these results along all the R2, VIF and p-values that, all together, highlight the appropriateness of the selection made.

5.3.1 The neural network prediction model

Following the above analysis, we constructed neural network models based on what the statistical analysis revealed as the best predictors. In this way, statistical analysis served as a first level of result interpretability, allowing us to highlight and gain insight into the inferences in place. After this, the prediction power of NNs was called upon to calculate prediction accuracies.

The Tensorflow machine learning library was used to assist this investigation. Neural networks were created with the above-discussed input and output layers, with two hidden layers in-between. The data were split in two parts; 70% were used as training data to develop the model and 30% as testing data to calculate its accuracy. The results and accuracies are summarised presented in the following table.

Final Consumption	Accuracy (1-MAPE)	RMSE
NL	91% (very good)	0.045
DE	81% (good)	0.126
ES	91% (very good)	0.063
SE	85% (good)	0.089
Electricity Consumption	**Accuracy (1-MAPE)**	**RMSE**
NL	89% (good)	0.114
DE	58% (poor)	0.411
ES	95% (very good)	0.045
SE	84% (good)	0.095
GR	93% (very good)	0.055
PT	91% (very good)	0.089

Table 3: Performance of neural network-based forecasting models for all 10 cases investigated (4 for final consumption and 6 for electricity).

5.3.2 The counterfactual approach

A possible important decision that could be supported by means of the above investigations would be:

'What is the best action that I should take in order to achieve an x% reduction of greenhouse emissions in the mid-term horizon?'

Such questions are typically addressed using counterfactual analysis. A counterfactual explanation of a prediction describes the smallest change to the feature values, which changes the prediction to a predefined output. **Counterfactual analysis is also referred to as local interpretability** in the sense that it does not aim to propose some general surrogate and more transparent model in the place of the typical black box of the neural network. Instead, **it aims at addressing 'what if' type questions and finding the minimal tweak of the model features that could secure this new goal.**

A first step towards interpretability is the selection of features via a statistical analysis, as shown above. This process allows us to gain insight into what really matters. An NN model would not provide any such service. **A next step for local interpretability would be to lay out a counterfactual analysis allowing us to address questions such as the one above. If we are to realistically tweak model features to perform 'what if' analyses, it is critical to identify the actionable features.** One cannot possibly change the weather by reducing the GDP/capita. In our case, the only possible actionable feature pertinent to our decisions here is that of energy taxes. Indeed, taxes in our analysis appeared in most cases as a key driver of consumption.

At this point, we should recall that there is an ongoing debate in the literature as to if and how much taxes can affect consumption. First, we have to acknowledge that not all societies respond in a similar way to taxes. Then there is always the possibility that there is a confounder to taxes; some other parameter that is truly causing the change, but as it moves in line with prices, one may end up with the wrong impression that it is prices that are driving consumption. Along this line of thought, a good example is provided by Borestein Severin[6], who is in favour of using the energy pricing mechanism. He argues: 'accounting for externalities requires introducing the 50 USD/ton CO2. The trend that taxes have shown convinces this will have an impact. Perhaps not direct—by immediate behavioural change—but by long-term driving for more innovation.' Indeed, prices bundle together three types of impact: the immediate behavioural response, a gradual behavioural change, and an impact on innovation. Perhaps the immediate response is not as strong, and perhaps this is why in the literature, there is often a claim for an essentially inelastic demand. However, how inelastic can demand be to price if it can trigger innovation or more mid-term behavioural shifts? Can we really claim that consumption is inelastic to prices if prices are driving innovation?

Below, we perform some counterfactual analysis on the results achieved via tweaking energy price/taxation. The table below illustrates the taxation change that would result in a 5% reduction of consumption in the seven cases overall, where taxation was found to be a driver of consumption. Both linear regression and NN models are reported.

a. *What-If Scenario for a 5% Decrease in Final Consumption via Taxation*

We present below the results of the taxation counterfactual analysis for both modelling approaches (linear regression, NN) for the case of final consumption . **Taxation appeared to be an important predictor in all four cases investigated for residential final consumption and electricity.**

Country	Most Recent Year Gas Tax (€/GJ)	Required Tax on Gas to Achieve a 5% Reduction on Final Consumption (€/GJ)		Increase Percentage (%)	
		Linear Regression	NN	Linear Regression	NN
NL	12.24	14.38	14.45	17.5	18.0
SE	14.206	24.48	25.85	72.3	82.0
ES	4.361	5.65	5.72	29.6	31.2
DE	4.404	7.41	8.2	68.4	86.4

\

Table 4. Results of the counterfactuals analysis on the taxation for both the linear regression and well as the NN models. The case of residential final consumption.

b. *What-If Scenario for a 5% Decrease in Electricity via Taxation*

Follow below the results of the taxation counterfactual analysis for both modelling approaches (linear regression, NN) for the case of electricity.

[6] Borestein Severin, Calculating the Effect of $50/tonne CO2 on Energy Prices, Energy Institute Blog, 2019, UC Berkeley. Available online: www.energypost.eu (accessed on 4 July 2021).

Country	Most Recent Year Tax on Electricity (€/KWH)	Required Tax on Electricity to Achieve a 5% Reduction on Electricity (€/KWH)		Increase Percentage (%)	
		Linear Regression	NN	Linear Regression	NN
ES	0.052	0.062	0.062	19.2	19.2
DE	0.159	0.268	failed	68.5	failed
PT	0.125	0.129	0.14	3.2	12

Table 5. Results of the counterfactuals analysis on the taxation for both the linear regression and the NN models. The case of residential electricity.

Above, we have restricted the analysis to households. However, imagine we could have similarly constructed models for the other two broad categories of energy consumption: transport and business/industry. In this case, our decision would also be informed by the other two models and would **require cross model counterfactuals, able to tweak all actionable features they may include, to find the least change and action required in order to achieve our end goal**, as articulated at the beginning of this section. It would be able to support energy policy decisions in a much more comprehensive way. One should also note that while the residential counterfactual presented above is, from a technical point of view, easy to elaborate and implement, this multi-model-ensemble-counterfactual would represent an AI challenge and will be a key area of investigation in the coming period.

5.3.3 Data and Model public Access (country case)

The residential energy datasets compiled for all investigated countries are available publicly online at

https://osf.io/vtzw8/

The files that can be found at this URL there are .jasp files and include besides the raw data itself also the JASP Statistical pre-processing (described above).

Models are and will remain publicly available and update continuously at

https://drive.google.com/drive/folders/1Y7mwyrW_4FsKnHHWMHhzmpBr7XPZ4-U1?usp=sharing

In the folder /COUNTRY ANALYSES

5.4 The next steps in the country/long-term sub case

Although the domain specific approach described above for the residential domain has some per se significance and can yield interesting and policy pertinent results, **the ability to treat all three domains, residential, transport and industry in one common framework and to run ensemble counterfactuals would represent a unique achievement.** There are a number of preliminary steps that would come with some independent value but would also create the conditions to culminate to this major development. Here is a summary highlighting the steps forward in this particular use case.

1. The statistical pre-processing is methodologically essential for all three domains and is an element of the overall explainability oriented approach; at the moment the residential and the industrial domains have been elaborated and have also been published. **The transport case is currently in elaboration and will complete some time in 2023.**
2. The NN analysis, illustrated above in the case on residences presents some tangible benefits that have been harnessed in terms of higher accuracies as shown above. Also, it is important that a counterfactual perspective has been feasible and has yielded interesting results. This said, **NN is not the best possible approach for explainability; GP and symbolic expression approaches, which are key in TRUST-AI, should be tested to see if some comparative advantage can result.**
3. The NN approach has been initially implemented in a proprietary environment (leiminte) and is not as transparent as it should ideally be. **We are completing the process of overhauling the discussion and replicating it in a more open environment (jupyter notebook) similarly to what has been done in the other subcase**. **The URL is provided above, in 5.3.1.** This is in line with the open data approach underpinning the use case and the project overall, and the will to publish fully transparent data and methods.
4. Last and most importantly, all above would allow **the implementation of an ensemble counterfactual** across the three domains, as presented in several above instances and illustrated in Figure 3, would represent a major advancement of the state of art in this thematic area. For this, we will need to complete all above points; in summary:

   - to complete point 1 above; transport analysis and models
   - to complete point 3 above; generate all models as open jupyter files
   - to complete point 2 above; develop open GP models preferably in the TRUST-AI framework so that we may make use of its expression pruning capabilities

6. Interaction and validation with business experts

As of late '21 we have established an **energy expert group**, comprising four experts. Recently this has been expanded with two more experts. The selection has been done on the following grounds:

- **deep awareness** of the building energy / ICT issues from a business point of view
- **close and tested business relationships**, ease of communication and a lack of any potential conflict of interest
- **potential for collaboration in the exploitation of results** (this applies particularly to the two 2022 experts)

The first interaction with the first batch of the 4 experts (referred to in detail in D 2.1) occurred via questionnaires and- especially- in depth, 1 hour online interviews, to which other TRUST AI partners also participated (Un. Tartu). The aim of this interaction was to collect general specification and orientation insights. The main takeaways from this interaction (reported also in D 2.1) was as follows:

- **Focus on the feature importance** and effort to convey this convincingly to the users, also by means of simple and effective user interfaces.
- **Focus on counterfactual analyses on actionable parameters** and again effort to convey this convincingly to the users, also by means of simple and effective user interfaces
- Additionally, **scepticism was expressed as to the potential of symbolic expressions to capture the phenomena**; the issue was considered interesting but the eventual possibility of such concise and meaningful expressions rather unlikely given the complex and nonlinear phenomena involved.

Towards mid' 22 the discussions were resumed with these same experts based on the results delivered at that moment in time. In fact, some first results on GP and symbolic expressions were made available, while the investigation on feature importance and counterfactuals had not yet started. This second round of discussions yielded the following feedback.

- **The issue of fast training was raised** as it was considered that monthly training data (the modelling till that moment in time has a training set spanning over 45 days). This was a highly value adding discussion raised by Mr. Alfio Galata but validated by all experts. Indeed for a practical demand response controller (that is the end product embedding all the modelling issues) such long training data would be a deficiency.
    - **The suggestion was made and the action was taken** to shorten the training time-frame (to 7- 10 days) and investigate the impact on accuracy. This does not prevent having more accurate models evolving in time. It would, however, allow the controller to become fastly operational, something of high deployment value. This proposal was immediately sent for implementation and some first results have been released even by the time of the first review, although the investigation is ongoing to reach definite conclusions.
- **The GP models and their symbolic expressions were discussed.** These models, for the first time even developed in the literature, surprisingly came with a higher accuracy (!) than the mainstream NN models used. However as suspected the expressions were not meaningful and did not add much to the user's insight.
    - **The suggestion was made and the action was taken** that- just as proposed from the first round of contacts (late '21)- the focus should shift to local/ instance explainability and not expect much from global model explainability and symbolic expressions. Dr. Hadjiyannis (whose business has activities in broad building AI) in particular suggested to look into the SHAP potential that was suggested and a model agnostic approach SHAP approach. This suggestion has been fully already uptake and it is currently the keyfocus of the investigation. This is illustrated also in the publication currently in preparation where the SHAP topic is highlighted even in the abstract.

https://docs.google.com/document/d/1YUDf9akHtitgCCOXbcQUov5gsOvZglpMZ6_oQFjpMUQ/edit?usp=sharing.

- However, **as APIN we also suggested not to discard but to to further investigate in the potential of symbolic expressions** especially as the TRUST AI framework would allow to easily carry on such investigations.

- **An important remark was raised with regard to the potential of explainability, and the encompassing demand response controller, in the residential context**. Indeed this suggestion was even raised in the first round of discussions so it was rather reiterated at this moment in time. Indeed the residential context is highly important and an excellent ground for demand response schemes. However, it is far more difficult than the office context, as energy uses are far more ad hoc and unpredictable; thus, the modelling approach needs to differ

  - **The response was given that the residential context will be investigated.** As APIN we have numerous related installations so the data collection would not require to repeat a yearly data collection. However, we are still uncertain if it is wise to diversify also in this context and not proceed full speed in the office context. Thus, although the suggestion is fully appreciated in its potential we will need to discuss this issue within the consortium to see what the modelling options could realistically be. In any case, given the data easily available, some trialling will in any case be attempted.

6.1 The future of expert interaction

A next round of expert discussions is planned for spring 2023 when

- **SHAP approaches will have been tested** and related results sent for publication
- more **insights will be available on symbolic expressions**
- a **concept for expanding to the residential concept will be available**
- a **first approach to counterfactuals will have been elaborated** and also tested either independently or even better in the TRUST AI

6.2 Expert validation on exploitation

As of late '22 we have established an early exploitation related contact with a major EU energy technology provider https://energy.ubitech.eu/.

The company has expressed strong interest in the broad approach and it was agreed to have a number of bilateral meetings (currently ongoing) to cross fertilise concepts. We have decided to develop pitching videos and presentations as a basis for clearly getting across what TRUST AI Energy is about and allow us to present it to broader exploitation oriented venues.

At this moment there are some important aspects not clarified in order to package and get across the business message; especially whether we will engage with the residential context (see discussions above), something that will largely affect the business plan.

The plan is to have this early business visibility in place by mid ‘23 at which moment we will leverage the ongoing discussion with the above mentioned provider.

6.3 Interacting with facility managers

We have had the opportunity to discuss some aspects of the work in TRUST AI with facility managers in buildings where we have data collection setups via our building management systems. The overall conclusion is that these users not only understand but intensely demand incorporation of demand response aspects in their setups. This provides some validation of the orientation of the project.

Yet, we have not yet raised explainability within this particular audience. Facility managers understand what demand response is, they also understand what the associated risk is. Now, the idea that explainability will mitigate this risk is not really understood.

Overall we plan to interact with experts from the facility manager community we have access to, but only **when a visible and tangible solution is there to show**. This is planned for early '24 when a fully fledged prototype is aimed at.

7. Conclusions

Below we discuss the next steps in the use case, as well as its integration within the TRUST-AI framework.

7.1 Roadmap and future developments in the Use Case

Follows below **a roadmap of the upcoming developments as regards the Energy Use Case.**

No	Month/ Year	Activity
THE SHORT TERM SUB CASE		
Modelling; short term		
1	02/ 2023	Elaborating of seasonal office building forecasting models & symbolic expressions (4)

2	02/ 2023	Experimenting with additional features (calendar days, etc.)
3	02/ 2023	Elaboration of counterfactuals/ explanations
Evaluation & Decisions; short term		
4	06/ 2023	Benchmarking of the GP against the NN models; accuracy and explanatory potential
5	06/ 2023	Local explainability; counterfactuals
6	06/ 2023	Global explainability; using TRUST AI framework to prune symbolic expressions derive building level/ seasonal/ yearly symbolic expressions and evaluating their global explainability potential
Integration; short term		
THE LONG TERM SUB CASE		
Modelling; long term		
7	12/2023	Derivation of NN/ GP models for transport and industry in our six countries
8	06/2024	Elaboration of ensemble counterfactuals for the country level forecasting
9	12/2024	TRUST AI framework integration; pull models/ explanations in production environment- country case

7.2 Linking to the TRUST-AI Framework

POLIS-21 is currently developing in Korea a demand response controller. Although this is independent of TRUST-AI elaborations, it can nevertheless serve as a quality testbed for testing the added value of explanations in energy related use cases. In addition, as this development runs quite in parallel to TRUST-AI the timelines are quite well aligned for the testing. The figure below illustrates the **four components of the solution and how they fit together.**

Follows a short description that will also **highlight the well defined link to the TRUST-AI framework.**

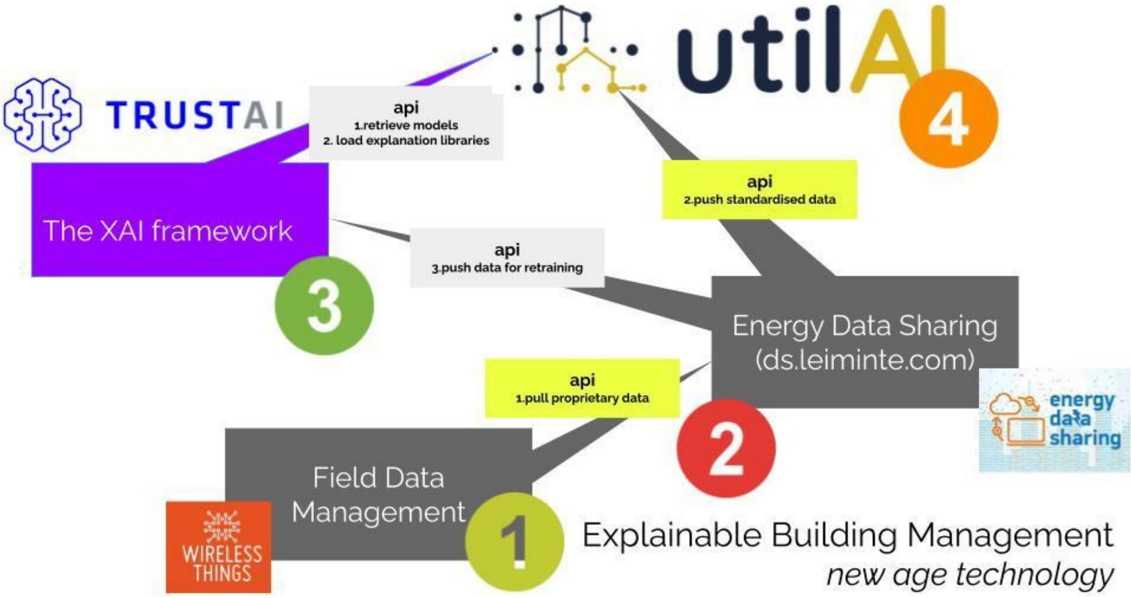


Figure 18: A true environment for testing the added value of GP models and explanations in energy apps

- **Component 1 (WT)** is the commercially available solution that has been used for real-time, field data sourcing in the pilot building
- **Component 2 (DS)** is a middleware solution developed in D 1.2 and accessible online at https://ds.leiminte.com. The purpose of this middleware is to provide for **open data access, via its documented API**. Thus, instead of feeding in the end solution (Component 4) proprietary data from Component 1 one will be able to export his data to Component 2 and then Component 4 will be able to read them. This is **an important contribution to open access approaches that will be heavily published and promoted as a key result**. Via this platform all the data used in the modelling will be shared with third parties who would like to use them in their developments.
- **Component 3 (TRUST-AI framework)** is the workbench where models and respective explanations will be developed. DS, being an open platform, will be able to export data to TRUST-AI as shown in the figure above.
- **Component 4 (util- AI)** is an in-process, in development (and unrelated to TRUST-AI solution) that looks forward to receiving models/ explanations from TRUST-AI and thus enhancing in value and functionality. Util-AI has been perceived as an environment running on an embedded and hardwired NN model. The ability to
    - **use TRUST-AI to support model flexibility** (instead of a hardwired NN model) and
    - **attach respective explanations**, along with what has been discussed above

  will significantly increase its value. UTIL-AI will increasingly be used for running models in the period ahead and hopefully interfacing with TRUST-AI, as well as DS.

- Elaborating (especially pruning and customising) symbolic expressions

and

- Exporting models and their associated explanations to a productive environment

are the two key points where TRUST-AI and the Energy Use Case connect and interact.

Indeed, the TRUST-AI framework is conceived in the work-plan as a standalone environment. APINTECH commits to undertake this additional interfacing work (pulling in data and pushing out models & explanations elaborated, provided

- the pertinence to the project and the added value is agreed by all
- IP rights issues are resolved

8. Recommendations

Below we list the three types of requirements that the Energy case would set to the TRUST AI framework.

Model & Explanation sourcing; The Energy instance requirements

A) Models' training (Framework service)
- manual data (clean) upload
- model training
- symbolic expression pruning
- model/ symbolic expression finalisation
- explanatory insight of symbolic expression assessment (Y/N)
- model/ symbolic expression saving
- model/ symbolic expression management
- classify as subtype (e.g. seasonal)
- delete model

B) Explanation development (Framework service)
- counterfactual definition
- counterfactual implementation
- explanations issued
- explanations saved

C) Model export to production environment (APIN)
- select model/ explanations
- export model/ explanations to production environment (UTIL AI)

Indeed, one can note that the Energy Use Case does not have any specific requirements as regards user interfaces as these will be separately implemented in the UTIL AI. UIs are of high importance and a generic approach can never be adequate when a business plan is in mind.

Reversely there are some aspects and functionalities for the framework that were not initially included in its specification (noted above under C) above). These are imperative for the Energy UC as **we plan to eventually use these models and explanations not in TRUST AI itself but in a production environment, UTIL AI** .

As APINTECH we have offered to implement this additional import/ export functionality provided it is considered of a broader utility to the other Use Cases and beyond.

9. Annex- open data and published work

Work till now has produced the following open publications. Additionally there is a significant amount of data and models that have been openly shared as disclosed above in 4.3.4 and 5.3.3 for the two sub use-cases respectively

9.1 Journals

1. Sakkas, N.; Yfanti, S.; Daskalakis, C.; Barbu, E.; Domnich, M. **Interpretable Forecasting of Energy Demand in the Residential Sector**. Energies 2021, 14, 6568. https://doi.org/10.3390/en14206568

   *Introduces the concept of **local interpretability/ counterfactuals analysis for the EU household sector**, based on data for 5 countries (DE, NL, PT, ES, GR).*

2. Sakkas, N., Yfanti, S. (2021). **Open data or open access? The case of building data**. Academia Letters,Article 3629. https://doi.org/10.20935/AL3629.

   *Discusses **open data issues and management as they pertain to the Energy Use Case**.*

3. Sakkas, N., Athanasiou, N. (2021). **Drivers of and counterfactuals for the final energy and electricity consumption in EU industry**. Academia Letters, Article 3451. https://doi.org/10.20935/AL3451

   *Introduces the concept of **local interpretability/ counterfactuals analysis for the EU industry** based on data for 5 countries (DE, NL, PT, ES, GR)*

9.2 Conferences

1. N. Sakkas, M. Papadopoulou, D. Sakkas, **Real time Data and Application Sharing and Collaboration for the Building Energy Domain**, World of Digital Built Environment WDBE 2021, access link

Discusses ***open data issues and management as they pertain to the Energy Use Case***.

9.3 Conference and Journal

1. Nikos Sakkas, Ch. Chaniotaki, Nikitas. Sakkas, Costas Daskalakis, **Building data models and data sharing. Purpose, approaches and a case study on explainable demand response**, Emerging Concepts for Sustainable Built Environment, Helsinki, November 2022

   Discusses the link between ***explainability considerations in TRUST AI and the requirements*** *of a demand response controller (UTIL AI)*