D4.2 - BEYOND Baseline Data Analytics – Draft Release

Grant Agreement n°	**957020**
Project Acronym	**BEYOND**
Project Title	**A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond**
Starting Date	01/12/2020
Duration	36
EU Project Officer	Stavros STAMATOUKOS
Project Coordinator	UBITECH
Consortium Partners	VTT, FVH, CIRCE, Suite5, IGM, KONCAR, ARTELYS, MYTILINEOS, CUERVA, BELIT, URBENER, BEOELEK,
Project Website	beyond-h2020.eu
Cordis	https://cordis.europa.eu/project/id/957020

Deliverable No.	**D4.2**
Deliverable Title	**BEYOND Baseline Data Analytics – Draft Release**
Work Package	WP4 - AI Big Data Analytics Toolkit with Data Sharing Functions
WP Leader	Suite5
Due Date	31/03/2022
Actual Date of submission	31/03/2022
Version	V1.00
Status	Final
Dissemination Level	Public
Authors	Suite5
Reviewers	UBITECH KONCAR

Disclaimer: *The present report reflects only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.*

Version	Modification(s)	Date	Author(s)
0.10	Draft ToC	01/02/2022	Suite5
0.20	Draft sections 1	03/02/2022	Suite5
0.30	Draft section 2	17/02/2022	Suite5
0.40	Updated sections 2	14/02/2022	Suite5
0.50	Updated sections 2, 3	24/02/2022	Suite5
0.60	Updated as per internal review	03/03/2022	Suite5
0.90	Final draft ready for peer review	18/03/2022	Suite5
0.91	Updated as peer review comments	30/03/2022	Suite5
1.00	Final Version submitted to EC	31/03/2022	Suite5

EXECUTIVE SUMMARY

D4.2 “BEYOND Baseline Data Analytics – Draft Release” reports on the preliminary activities undertaken in the context of T4.2 “Data Analytics Algorithms Baseline Definition” and T4.3 “Scalable, Secure Multi-Party Computations and Edge Analytics Execution Engines” of WP4 “AI Big Data Analytics Toolkit with Data Sharing Functions” by M16 of the project implementation. The main scope of D4.2 is to present the draft release of the BEYOND baseline data analytics functions, offered to the users by the BEYOND Platform’s Isolated Data Analytics Containers which incorporate various data analytics functions (such as pre-trained models and algorithms) for usage in personal and industrial data analysis.

As the actual deliverable is of type OTHER, this document comes as an accompanying report summarising the overall implementation of the developed data analytics functions and the associated implementation mechanisms, namely the PreTrained Analytics Sets, which are available in a dedicated online wiki containing the full content of each analytics solution.

It shall be noted that, the context of the BEYOND Pretrained Analytics Sets will be continuously updated and enriched to incorporate all planned and new solutions that may appear as the project’s activities evolve, especially during the development activities of the various BEYOND End-User tools in WP5 and WP6 and during the demonstration activities in carried out in WP7, where further analytics requirements is anticipated to emerge.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

Abbreviation	Full text
AI	Artificial Intelligence
CIM	Common Information Model
DER	Distributed Energy Resource
DL	Deep Learning
Dx.y	Deliverable x.y
DoA	Description of Action
IDC	Isolated Data Containers
ML	Machine Learning
Mx	Month x
Tx.y	Task x.y
UC	Use Case(s)
UI	User Interface
WPx	Work Package x

1. INTRODUCTION

1.1 Scope of the document

The focus in D4.2, is given on the definition of a set of pre-selected and pre-trained data analytics algorithms and models and their consequent population in the BEYOND Cloud based Platform's Pretrained Data Analytics Sets, which essentially represent the data analytics catalogue offered by the core platform.

In alignment with the DoA, the initial activities focused on identifying existing state-of-the art implementations, existing libraries and algorithms for knowledge extraction and business intelligence on data analytics, along with identification of appropriate visualisation libraries that can be used to effectively communicate the outcomes to the BEYOND building-related and energy stakeholders.

The pre-trained data analytics offered through the BEYOND platform will span over 3 baseline dimensions: (a) personal data analytics (e.g. for consumer energy behaviour, comfort preferences and flexibility), (b) industrial data analytics (indicatively involving Building Energy and Predictive Maintenance analytics, DER Forecasting and Flexibility analytics, High-level building portfolio analytics) and (c) real- time edge analytics (e.g. for flexibility services related to Smart Home/ Building Automation).

It shall be noted that the list of the BEYOND PreTrained Analytics Sets (presented in D4.2) will be further enriched to include all the planned analytics solutions, as well as any new ones that may emerge during the project's implementation. The developed analytics will be continuously upgraded and improved to fulfil the BEYOND stakeholders needs and further requirements that are anticipated to emerge during the development activities of the BEYOND End-users tools in WP5-WP6 and consequently in the WP7 demonstration activities.

1.2 Relation to other tasks/deliverables

The present deliverable documents the activities performed in the context of T4.2 and T4.3 by M16 of the project's implementation and its main scope is to deliver the draft version of the BEYOND Pretrained Analytics Sets, i.e., an analytics catalogue containing pre-trained analytics models and algorithms which can be used by authorised users of the BEYOND platform, towards extracting new knowledge and intelligence on their data assets.

Towards this direction, D4.2 receives input from the following BEYOND tasks and associated deliverables:

- T2.5 “Detailed architecture design, protocols, and interfaces specifications for Big Data-enabled Energy Services”; where in D2.6 the initial design specifications for the Data Analytics services bundles are documented.
- T3.3 “Platform Backbone Infrastructure, On-Premise and Secure Experimentation Playground Data Containers and Core Services Development”
- T4.1 “Big Data Analytics Workbench and Jobs Execution Engines”

The outcome of the activities performed in D4.2 will be also used as input in the following BEYOND deliverables and work packages:

- D4.3 “BEYOND Data Analytics, Sharing & Matchmaking Services Bundles – Release 1.00”
- D4.4 “BEYOND AI Data Analytics Toolkit – Final Release”

Moreover, D4.2 offers a better understanding of the different data asset analytics functions offered by the BEYOND Platform to all BEYOND End-users tools that will be delivered in WP5 “AI Analytics-based Decision Support Suite for Optimizing Energy Policy Planning, Infrastructure Sizing and De-risking Renovation Investments and in WP6 “AI Analytics-based Innovative Energy Services Suite towards Optimized Buildings Energy Performance Management”.

1.3 Structure of the document

Towards addressing all the aspects relevant to the scope of D4.2, the remaining of this document has been structured as follows:

- Chapter 2 presents the rationale and objectives of the BEYOND PreTrained Analytics Sets, and its draft version
- Chapter 3 presents the conclusions and the next steps towards the final release of the PreTrained Analytics Sets.

2. BEYOND Baseline Data Analytics

2.1 Rational and objectives

The BEYOND solution aims to deliver to its various building and energy-related stakeholders, valuable insights and appropriate visualisations based on various pre-trained analytics models and decision-making algorithms that can be applied onto the data assets available in the BEYOND Platform and deliver added value to their business operations.

As described in detail in D2.6 and in D4.1, this functionality is facilitated by the BEYOND platform's **Isolated Data Analytics Container**, offering to its users a configurable framework, namely the **Analytics Composer** for designing analytics workflows using (once authorised) the data assets residing in the platform, towards receiving readily available insights, that can address their needs. This component is heavily supported by the **PreTrained Analytics Sets** which essentially represents the data analytics catalogue of the BEYOND platform, offering various pre-trained analytics models, algorithms and pre-configured analytics workflows to facilitate the analytics design and reduce the configuration and execution time of the analytics algorithms.

Overall, the BEYOND PreTrained Analytics Sets subscriptions range from statistics computations and visualisations to pre-trained machine learning (ML) algorithms (e.g., regression, classification, clustering, forecasting, etc.) deep learning (DL) models, well-developed analytics workflows, and configurable model training workflows, which aim to enhance the overall decision-making capabilities of the BEYOND stakeholders.

Clearly, this component has a key role in the overall BEYOND solution, offering to its users various analytics functionalities that address specific needs of the domain in a flexible manner; thus providing readily available insights into common domain issues, also accelerating the development of building energy related solutions, targeting the exact stakeholders' needs.

The analytics functions offered through the PreTrained Analytics Sets component delivered in this draft release, are based on the BEYOND Use Cases and functional requirements as defined in D2.6 and is anticipated to be further updated and enriched to include additional functions addressing the needs of the BEYOND End-users tools developed in the context of WP5 and WP6 activities and address any further needs that may emerge during the demonstration activities of WP7.

2.2 Documentation methodology

As the context of the BEYOND analytics catalogue (i.e., the PreTrained Data Analytics Sets) is anticipated to evolve continuously throughout the project' implementation, it

has been agreed among the partners to utilise a dedicated wiki page: https://wiki.beyond-platform.eu in order to record and describe all ongoing work; acting as the live documentation for the BEYOND PreTrained Analytics Sets. A screenshot of the respective wiki is shown in the following figure.

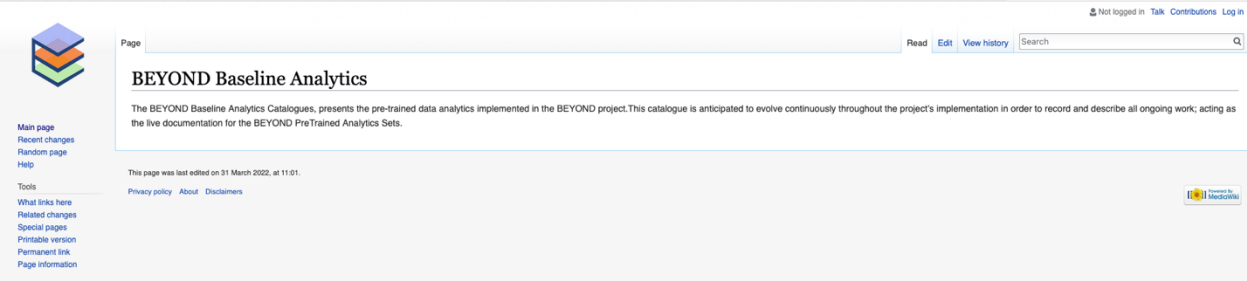

FIGURE 1 BEYOND ANALYTICS WIKI

To capture all necessary information in a consistent manner, a template has been shared among the partners to fully describe their offered analytics solution. A brief description of the content covered is described as follows:

1. **Rationale & Link to BEYOND Apps**; here the problem that is addressed through the developed solution is described also presenting how the proposed analytics solution can be leveraged through BEYOND.
2. **Overview of relevant implementations**; here the material /knowledge used to better comprehend the problem and effectively design the proposed solution is presented, including relevant state of the art, a brief landscape analysis of existing research related to the problem's solutions and identification of similar implementations
3. **Implementation in BEYOND**
   a. ***Data inputs and Analytics Pipeline (incl. assumptions /limitations)***, here the data used as input to the model are presented (both for baselining and training). In addition, a high-level conceptual description of the analytics workflow(s), is provided, including any required pre-processing steps for the proposed solution; the leveraged technologies, along with any assumptions taken during the model development and any identified limitations.
   b. Analytics Libraries Employed, here the technologies leveraged for data manipulation are presented, along with the data analytics libraries used in the proposed model.

2.3 BEYOND Pretrained Analytics Sets (Draft release)

In alignment with the DoA, the pre-trained data analytics in BEYOND span over three baseline dimensions:

1. *Industrial data analytics*, aiming to address the various buildings' energy-related issues at a high level, involving Building Energy and Predictive Maintenance analytics, Demand and Generation Forecasting and Flexibility forecasting.
2. *Personal data* analytics which aim to address the various building energy-related issues (at lower level) from the occupant's viewpoint involving occupant's energy behaviour profiling and forecasting, comfort preferences profiling and flexibility.
3. *Real-time edge analytics*, which aim to enhance a building's "intelligence" through automated control and optimisation such as flexibility services related to Smart Home/ Building Automation.

It shall be noted that in this draft release only the first two dimensions are considered, while the real-time edge analytics will be addressed in the upcoming final release of the BEYOND Baseline Data Analytics (due in M34 and documented in D4.4).

Based on the nature of the underlying problem an initial categorisation of the Industrial and Personal data analytics issues that are addressed from the BEYOND Data Analytics services is presented in the following schematic and described as follows

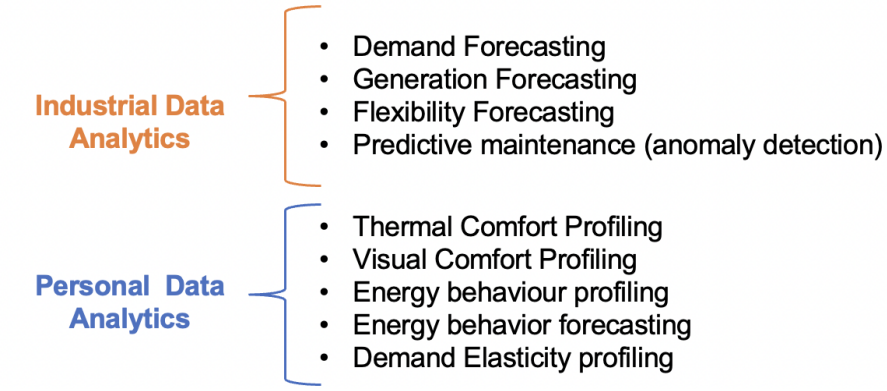

FIGURE 2 BEYOND DATA ANALYTICS CATEGORISATION

In the case of **Industrial** data analytics, a) *Demand Forecasting*, refers to problems of energy demand forecasting at Distributed Energy Resource (DER) level b) *Generation Forecasting*, refers to problems of energy generation forecasting at DER level, c) *Flexibility Forecasting*, refers to identifying the share of demand/generation that can be reduced, increased, or shifted within a specific duration and d) *Predictive Maintenance,* refers to proactively identifying (through anomaly/outlier detection) when maintenance on building energy-related devices (e.g. HVAC, Lighting, etc) is required.

Similarly, in the case of **Personal** data analytics, a) *Thermal* and b) *Visual Comfort Profiling* refer to problems associated with building occupants' thermal and visual comfort preferences respectively; c) *Energy behaviour profiling* refers to problems around building occupants' energy consumption behaviour and d) Demand Elasticity refers to the elasticity of demand

The pre-trained models/analytics have been prioritised based on their relevance to the business scope of the project, the respective WP4 requirements, the preliminary analytics needs from the BEYOND End-user Tools (WP5 and WP6). The collective list also presenting their category and their prioritisation by denoting the respective deliverable (D4.2 or D4.4) in which they will be incorporated, is provided in the following table.

TABLE 1 BEYOND PRETRAINED ANALYTICS SETS

#	Category	Title	Release
1	**Personal Analytics**	Thermal Comfort Profiling (Boundaries)	D4.2
2		Thermal Comfort and Duration of retaining different setpoints within comfort boundary conditions	D4.2
3		Visual Comfort Profiling (Boundaries)	D4.4
4		Visual Comfort and Duration of retaining different setpoints within comfort boundary conditions	D4.4
5		Prediction for the use of devices in the next hours (and which devices) for an occupant	D4.4
6		Behavioural Clustering of Occupants	D4.4
7		Demand forecasting (at device level) under different comfort levels	D4.4
8		Demand elasticity profiling per consumer	D4.4
9	**Industrial Analytics**	Short-term demand forecasting at building/ portfolio level	D4.2
		Long-term demand forecasting (building/ portfolio) / electricity and heating	D4.2
		Very long-term demand forecasting (building/ portfolio) / electricity and heating	D4.2
		Generation forecasting DER level in short (6-24 hours), very short term (15 minutes-1-hour ahead)	D4.2
		Mid-term generation forecasting at DER level	D4.2
		Very-long term generation forecasting at DER level	D4.2
		Prediction of demand flexibility at device level - short-term (6-24 hours ahead?), very short term (15 minutes-1 hour ahead?)	D4.2
		Anomaly/outlier detection in the energy performance of HVAC system components (e.g., heat valves, filters, fans, etc)	D4.2
10		Demand forecasting (at device level) under different comfort levels	D4.4
11		Very Short-term demand forecasting at building/ portfolio level	D4.4
12		Very-short heating demand forecasting at building level	D4.4
13		Short-term heating demand forecasting at building level	D4.4
14		Mid-to-Long-term demand forecasting (building/ portfolio) / electricity and heating	D4.4
17		Anomaly/outlier detection in energy demand	D4.4
20		Long-term generation forecasting at DER level	D4.4
22		Anomaly/outlier detection in energy generation	D4.4
25		Clustering of consumers based on their demand elasticity	D4.4

It shall be noted be noted that the draft version of the analytics solutions forming part of the Pretrained Analytics Sets (i.e., baseline analytics portfolio) will be available in the BEYOND Cloud based platform's Marketplace, based on their defined license and respective access policies in the 1st stable release of the BEYOND Integrated Platform (Release 1.00) due on M20.

3. Conclusions and plan for final release.

D4.2 “BEYOND Baseline Data Analytics – Draft Release” presents the overall purpose of the BEYOND Pretrained Analytics Sets, being in essence the data analytics catalogue offered through the BEYOND Platform, incorporating industrial and personal data analytics (in this draft release) targeting the building energy related data value chain and which can be leveraged through the platform as preconfigure blocks for more advanced analytics services.

In D4.2, ten different Pretrained models are delivered as knowledge provided in the dedicated wiki platform, which will be transported to the platform and be deployed as ready-to-use solutions by the different stakeholders. These analytics solutions will be enhanced and adapted (if required) to address the requirements of the BEYOND stakeholders, anticipated to arise during the BEYOND End-user’s Tools development activities in WP5-WP6 and during the demonstration activities in WP7.

The updated version of this deliverable, namely D4.4” BEYOND AI Data Analytics Toolkit – Final Release” due in M34, will include the rest of the planned analytics functions (not prioritised in this release) along with any new ones that might arise during the development activities for both the BEYOND platform’s Isolated Data Analytics Containers and the Data Analytics Containers of the BEYOND Private Infrastructure.,

References

[1] BEYOND (2020) Description of Action (DoA)
[2] BEYOND (2021a): D2.1 - End-user & Business requirements analysis for big data-driven innovative energy services & ecosystems -a
[3] BEYOND (2021b): D2.6 - BEYOND Framework Architecture including functional, technical and communication specifications – a
[4] BEYOND (2022a): D3.3 - Data Collection, Security, Storage, Governance & Management Services Bundles – Beta Release
[5] BEYOND (2022a): D3.4 - BEYOND Integrated Platform –- Alpha, Mock-ups Release
[6] BEYOND (2022a): D4.1 - BEYOND Data Analytics, Sharing & Matchmaking Services Bundles – Beta Release