



Customer Digital Twin

D6.3

April 2022

Deliverable

PROJECT ACRONYM	GRANT AGREEMENT #	PROJECT TITLE
TWINERGY	957736	Intelligent interconnection of prosumers in positive energy communities with twins of things for digital energy markets

DELIVERABLE REFERENCE NUMBER AND TITLE

D6.3

Customer Digital Twin

Revision: v1.0

AUTHORS

Giannis Gialelis	Dimitris Karadimas	Chris Mountzouris	Tasos Kladas
UoP	UoP	UoP	UoP



Funded by the Horizon 2020 programme of the European Union
Grant Agreement No 957736

DISSEMINATION LEVEL

✓ **P Public**

CO Confidential, only for members of the consortium and the Commission Services

Version History

REVISION	DATE	AUTHOR	ORG...	DESCRIPTION
V0.1	28/02/2022	John Gialelis, Chris Mountzouris	UoP	First Draft for Review by QCG
V0.2	28/03/2022	John Gialelis, Chris Mountzouris Dimitris Karadimas	UoP	Updated draft to incorporate feedback and comments from the QCG
V0.3	26/04/2022	John Gialelis, Chris Mountzouris Dimitris Karadimas Anastasios Kladas	UoP	Inclusion of results from experimental tests for the purposes of validation
V0.4	26/04/2022	John Gialelis, Chris Mountzouris	UoP	Inclusion of results from experimental tests for the purposes of validation
V1.0	29/04/2022	Athanasios Chassiakos, Stylios Karatzas, Anastasios Karameros	UoP	Final review – Deliverable submitted to the EC

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both.

Legal disclaimer

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The above referenced authors shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the CINEA nor the European Commission is responsible for any use that may be made of the information contained therein.

© 2022 by TwinERGY Consortium

Executive Summary

The deliverable **D6.3: Consumer Digital Twin (CDT)** is implemented under the framework of **T6.3: Consumer Digital Twin (CDT) design & development** within the **Work Package 6: Development of Digital Twin Platform & System dynamics**. It describes the steps of the methodology framework which was adopted at the development process by CDT and the tasks included in each step. Moreover, it presents the front-end and back-end technologies which are implemented for the development of the CDT as a web app, as well as, the technologies utilized for data acquisition, data visualization and data analysis. A detailed description for the implementation of Keycloak authentication mechanism is provided, too. Finally, it presents algorithms and methods that supports the core functionalities of the CDT such as data-driven forecasts and decision support.

The current document is divided into four different sections that can be summarized as follows:

- Section 1 introduces the scope of this document as well as its contents.
- Section 2 introduces the technology of DTs and describes the differences between a DT and a CDT.
- Section 3 describes the methodology used for the development of the CDT, the CDT platform technologies, the authentication mechanism of CDT and presents the core graphical user interfaces of the CDT.
- Section 4 describes the algorithms and methods utilized by the CDT for the data processing and modelling.

Index

Legal disclaimer	3
Index	5
List of figures.....	7
List of tables	8
List of equations	9
1. Introduction	10
1.1 Purpose of this deliverable	11
1.2 Scope of this deliverable	11
1.3 Structure of the document	12
1.4 Abbreviation list	14
1.5 Audience	15
2. Analysis Study for DT and CDT Systems	16
2.1 Digital Twins; Definition.....	16
2.2 Enabling Technologies of Digital Twins.....	17
2.3 Model-Driven and Data-Driven Digital Twins.....	17
2.4 Digital Twins; Applications and Research.....	18
2.5 Differences between DT and CDT	19
2.6 Consumer Digital Twin; 3-tier Architecture	19
3. CDT Platform, Technologies, Data Aggregation, Standards and Interfaces.....	21
3.1 CRISP – DM Methodology	21
3.2 CDT Platform Technologies.....	23
3.3 Authentication through Keycloak.....	28
3.4 Interoperability (iSCAN)	36
3.5 Graphical User Interfaces & Dashboard	36
4. CDT Data Processing	41
4.1 Consumption Profiling	41
4.2 Production Profiling	46

4.3 Preferences; Determination of Importance	48
Annexes	64
A. CDT - Web-application Map	64

List of figures

Figure 1 The 3-tier Architecture of CDT	20
Figure 2. The steps of CRISP-DM Methodology.....	22
Figure 3. Keycloak identity flow explained with or without an external Identity Provider	28
Figure 4. Example SSO login page with Keycloak.....	29
Figure 5. TwinERGY Realm	30
Figure 6. TwinERGY client.....	30
Figure 7. Sample list of user attributes	31
Figure 8. Customization of user-defined attribute for TwinERGY scope	31
Figure 9. Assignment of a user-defined attribute for a user.....	32
Figure 10. Keycloak user roles	32
Figure 11. Role's mappings	32
Figure 12. Request cookies of a Keycloak protected TwinERGY resource	35
Figure 13. Example JWT response	36
Figure 14. Information about the install Solar PV within the occupancy.....	37
Figure 15. Aggregated daily and 7-day forecasted energy production	37
Figure 16. Environmental Conditions that affect the energy production level (1/2).....	38
Figure 17. Environmental Conditions that affect the energy production level (2/2).....	38
Figure 18. Dashboard of the CDT [1/3]	39
Figure 19. Dashboard of the CDT [2/3]	39
Figure 20. Dashboard of the CDT [3/3]	40
Figure 21. The notifications component of CDT	40
Figure 22. Forecast of Energy Demand with Seasonality Trend	46
Figure 23. The GUI for Preferences (1/2)	49
Figure 24. The GUI for Preferences (2/2)	49
Figure 25. Importance scale for Preferences	51

List of tables

Table 1. Abbreviation list.....	15
Table 2. Data-Driven DTs vs. Model-Driven DTs.	18
Table 3. A typical dataframe of Consumers and their collected features and characteristics	42
Table 4. Reduced dataframe as produced by PCA method	43
Table 5. Consumer’s Energy Demand for the past two years.....	45
Table 6. Forecast of Energy Demand with Seasonality Trend	46
Table 7. The required data for the forecast model.....	47
Table 8. The Preferences of CDT	50
Table 9. The point-scales used to establish the priority of criteria	51
Table 10. The importance level of each Preference for the Case Study	52
Table 11. The weight of each Preference as determined from the EQ Method.....	53
Table 12. The weight of each Preference as determined from the RI Method	53
Table 13. The weight of each Preference as determined from the ROC Method	54
Table 14. The weight of each Preference as determined from the Simos Method	56
Table 15. The definition and explanation scale values.....	57
Table 17. Decision Matrix of AHP Method.....	58
Table 18. The values of RCI for each n-value.....	59
Table 19. The decision matrix of VIKOR’s Method	60

List of equations

Equation 1. Linear Regression model	45
Equation 2. The equation that describes the energy demand	45
Equation 3. Equation that describes the T_{PV} when the value of wind speed is available .	47
Equation 4. Equation that describes the T_{PV} when the value of wind speed is not available	47
Equation 5. Equation that describes the P_{DC}	48
Equation 6. Equation that describes the P_{AC}	48
Equation 7. Determination of each Preference weight through Equal Weights Method .	52
Equation 8. Determination of each Preference weight through Relative Importance Method	53
Equation 9. Determination of each Preference weight through Rank Order Centroid Method	54
Equation 10. Determination of each Preference weight through AHP Method.....	58
Equation 11. Determination of Consistency Ratio (CR)	58
Equation 12. Determination of each Consistency Index (CI).....	59
Equation 13. The L^P -metric	60
Equation 14. Best-case and worst-case scenario determination for benefit criteria	61
Equation 15. Best-case and worst-case scenario determination for cost criteria	61
Equation 16. The S-metric.....	61
Equation 17. The R-metric.....	61
Equation 18. The Q-metric	62
Equation 19. The required condition to meet acceptable advantage.....	62
Equation 20. The required condition to meet acceptable stability	63

1. Introduction

TwinERGY introduces a first-of-a-kind Digital Twin framework that incorporates the required intelligence for optimizing demand response at the local level without compromising the well-being of consumers and their daily schedules and operations. The main idea behind the conception of the TwinERGY project lies on the interest of the project partners to exploit the new business opportunities that project implementation delivers and increase the relevance of the DR optimization tools and strategies in the new generation of energy management systems.

By coupling mature practice for citizen engagement with service innovation through the lenses of public value, TwinERGY will ensure that a wide range of interests and especially of consumers/prosumers will be represented and supported in the energy marketplace. In this context, TwinERGY will develop, configure and integrate an innovative suite of tools, services and applications for consumers, enabling increase of awareness and knowledge about consumption patterns, energy behaviours, generation/ demand forecasts and increase of local intelligence via properly established Digital Twin-based Consumer-Centric Energy Management and Control Decision Support mechanisms that locally optimize demand response. Key use cases will be trialed across 4 pilot regions making use of cutting-edge methods and tools. Special focus is given on standardization and policy & market reform as key enablers for the successful commercialization of the TwinERGY results. Additional attention is given in establishing knowledge transfer and exchange synergies with similar projects listed under the BRIDGE Initiative, while explicit focus will be given on the establishment of close collaboration with the projects funded under the LC-SC3-ES-5-2018 topic, to further reinforce and catalyze collaborative advancements in research, innovation, regulatory and market issues around Demand Response, RES Integration and Consumer Engagement.

1.1 Purpose of this deliverable

The deliverable D6.3: Consumer Digital Twin is implemented in the framework of T6.3: Consumer Digital Twin design and development within the Work Package 6: Development of Digital Twin Platform & System dynamics. It outlines the Consumer Digital Twin (CDT) which is an evolving digital profile depicting the past and current energy consumption behaviour of a household consumer. Its objective is the design and the development of the CDT aiming at one hand for the aggregation of real-time measurements and data such as energy consumption and production recordings, personal physiological data recordings, various tolerance and flexibility levels (comfort, eco, load, financial), indoor and outdoor environmental conditions, equipment properties, and the other hand for the optimization of energy consumption across an array of dimensions (domiciles, buildings, communities). Moreover, CDT's architectural aspects such as internal structure (meta models, standards), means to receive info (APIs), binding mechanisms, run time environment platform, realization features, use cases and GUIs are identified.

1.2 Scope of this deliverable

The scope of this deliverable is to depict the design and development of a data driven CDT, which was devised to monitor and control the past and current energy consumption behaviour (profile) of a domicile occupant, renewable energy production levels of the domicile, related indoor environmental and air quality conditions, outdoor environmental conditions, occupant's indoor thermal comfort status as well as occupant's thermal tolerance, load flexibility and financial ability. A detailed description will be provided of the corresponding data frames generated from the data collected via the dwelling's sensor network along with the tenant's manual inputs. The CDT modules / components together with their associated outputs will be presented and described in detail. Standardized APIs for the seamless binding and mechanisms for the automated participation in energy communities shall also be introduced. Moreover, the most essential characteristics of a CDT which are its high fidelity in terms of functionality,

appearance, and control combined with a high level of accuracy and reliability for simulations and forecasts will be highlighted.

1.3 Structure of the document

- **Analysis Study for DTs and CDT.**

This section provides an analysis of the current technological status of DTs and CDT, and provides insights regarding the continuous aggregation of data, the assessment of information, the knowledge generation, the optimal scheduling of the devices and equipment. A definition of the concept, vision and approaches of digital twin and consumer digital twin, objectives, conceptual models, main applications, differences, and research shall be presented. CDT architecture, implementation approach, proposed models as well as issues related to reliability of CDT are depicted.

This section is structured as showing below:

- **DTs**
 - The Digital Twin Vision
 - Model- and Data-based Predictions Systems
 - Data-Driven Approaches to Create Digital Twins
 - Executable Digital Twins
 - Main applications and research
- **CDTs**
 - Concept of Consumer Digital Twin
 - Augmented digital twin
 - CDT conceptual model
 - Differences between DT and CDT
 - CDT Architecture and Implementation approach
- **CDT Platform, Technologies, Data Aggregation, Standards and Interfaces**

CDT is an evolving data driven digital profile depicting the past and current energy consumption behaviour of a domicile occupant associated with any energy production

potential, outdoor and indoor environmental conditions, indoor air quality, occupant's thermal comfort and wellbeing status and occupant's preferences. In the CDT, there is an automatic bidirectional flow of data between the physical and digital objects and the changes in either object, physical or digital, directly lead to changes in the other. The data driven CDT is implemented following the six (6) steps of the CRISP- DM methodology. Data Aggregation (CDMT), Authentication (Keycloak), interoperability (iSCAN), CDT technologies and CDT GUIs data are displayed on a front-end dashboard which serves as an information management tool. This section is structured as showing below:

- **CRISP – DM Methodology**
 - CDT: Business Understanding
 - CDT: Data Understanding
 - CDT: Data Preparation
 - CDT: Modelling
 - CDT: Evaluation
 - CDT: Deployment
- **CDT Platform Technologies**
 - Functional and non-Functional System Specifications
 - Frontend
 - Backend - Server
- Authentication
- Interoperability
- Graphical User Interfaces
 - Graphical User Interfaces
 - Dashboard

Numerous studies have already demonstrated the role of Digital Twin as a front-runner in the virtual representation of physical assets. In the CDT, consumer preferences and IoT-enabled devices form the basis for a knowledge base environment in which personalized information can be deducted for each end user. The potential impact of CDT

on energy related services for the end user is expected to be substantial, by utilizing CDT's information regarding consumer's energy consumption and flexibility profile, thermal comfort tolerance, financial preferences, and domicile structure characteristics. The binding mechanisms for the automated participation in energy communities is expected to improve consumer's energy efficiency and price discovery. This section is structured as showing below:

- Models
 - Consumption Profiling
 - k-means clustering
 - Linear Regression
 - Production Profiling (rooftop PV energy estimation)
 - Preferences (thermal tolerance (static limits), wellbeing status significance, energy flexibility in hours, eco friendliness, financial choices)
 - Implementation of methods to determine the weight of each Preference.
 - Implementation of Vikor Method to support MCD Analysis based on Preferences.

1.4 Abbreviation list

AI	Artificial Intelligence
CDT	Consumer Digital Twin
CI	Consistency Index
CR	Consistency Ratio
CRISP-DM	Cross-Industry Standard Process model for Data Mining
DA / KD	Data Analysis / Knowledge Discovery

DT	Digital Twin
HVAC	Heating Ventilation and Air Conditioning
IoT	Internet of Things
LEM	Local Energy Market
MCDA	Multi-criteria Decision Analysis
ML	Machine Learning
TVOCs	Total Volatile Organic Compounds

Table 1. Abbreviation list

1.5 Audience

The present document (along with its updated versions) is a public deliverable, which is addressed to the TwinERGY consortium partners, the European Commission services and everyone interested in the rules and procedures of data management in TwinERGY project and will be published through all the appropriate channels of the project.

2. Analysis Study for DT and CDT Systems

2.1 Digital Twins; Definition

DTs are simplified virtual representations of physical entities or real-life systems within virtual environments. They comprise selected characteristics, properties and conditions of them¹. DTs aim to model behaviours or processes, execute data-driven or model-driven simulations, support decisions and make forecasts or predictions. DTs differ from other virtual representation concepts, such as Digital Shadows and Digital Models, as they support automatic bidirectional transfer and maintain high level of synchronization between the physical and the virtual entity².

The concept of DTs is not new. The term 'Digital Twin' was first defined by NASA in respect to the Aerospace Industry. NASA defined DT as "an integrated multi-physics, multi-scale, probabilistic simulation of [...] a system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding [...] twin". Since then, it has been described with various terms in the literature³. This permutation is the result of the DTs' utilization by various industrial fields, such as Manufacture, Energy, Transportation, Telecommunication and Health industries. A recent definition of DTs has been proposed by Stark and Damerou. It defines DTs as "a digital representation of an active unique product [...] or unique product-service system [...] that comprises its selected characteristics, properties, conditions, and behaviours by means of models, information, and data within a single or even across multiple life cycle phases."

¹ M. Singh et.al, Digital Twin: Origin to Future, Applied System Innovation, MDPI, 2021.

² D. Hartmann and H. Van der Auweraer, Digital Twins, Chapter 1, Siemens Lab.

³ D. Jones et.al, Characterising the Digital Twin: A systematic literature review, CIRP Journal of Manufacturing Science and Technology, Elsevier, 2020

Nowadays, the concept of DTs has been extended from non-living entities to complex bio-entities, such as humans⁴. Humans are modelled in respect to a specific aspect of their life, such as consumers within a market environment or patients within a medical environment. However, the Human DTs are still in the embryonic stem.

2.2 Enabling Technologies of Digital Twins

DTs employ various technologies at their development phase and operation lifecycle which allow them to represent physical entities within a virtual space, acquire near real-time data from multiple resources, handle big volume of data, process and analyze data, model behaviors and visualize information to end-users⁵. In respect to the sector of appliances, DTs utilize different technologies and tools. In any case, a DT should embrace at least three (3) technologies; IoT, AI/ML and Cloud Computing.

IoT is a core technology utilized by DTs, as it provides the capability of data capture from physical entities. Sensors embedded in smart devices allow data acquisition from multiple resources at the same time. AI/ML supports the core functionalities of a DT. They provide advanced tools, techniques and methods to analyze complex, multivariable and heterogeneous datasets. The data analysis produces valuable outcomes such as forecasts and data classification. Cloud Computing provides an efficient infrastructure for rapid and reliable access to data, as well as, a deployment environment for a DT.

2.3 Model-Driven and Data-Driven Digital Twins

DTs can be classified into two main categories due to the orientation of the implementation approach; **Model-driven DTs** and **Data-driven DTs**⁶. **Model-driven DTs** employ mathematical laws and physical models to execute simulations. They require deep domain knowledge of the physical entity and small volume of data. DTs that

⁴ S. Wei, Is Human Digital Twin Possible? Computer Methods and Programs in Biomedicine Update, Elsevier, 2021

⁵ M. Liu et.al, Review of digital twin about concepts, technologies, and industrial applications, Journal of Manufacturing Systems, Elsevier, 2021

⁶ R. He, Data-driven digital twin technology for optimized control in process systems, ISA Transactions, Elsevier, 2019.

represent non-viable entities which are characterized by low complexity, small volume of related data and non-multiparametric environments follow the Model-driven approach.

Data-driven DTs employ AI/ML methods and techniques to process raw data from the physical entity's environment and produce knowledge such as forecasts, pattern discovery and data classification. They require multivariable data from multiple resources and minimal domain knowledge. DTs that represent viable entities, such as humans, follow the Data-driven approach, as these entities cannot be described deterministically due to behavioural subjectivity.

Data-Driven DT	Model-Driven DT
Little domain knowledge	Deep domain knowledge
Big volume of data	Small volume of data
Forecasts, classifications	Simulations

Table 2. Data-Driven DTs vs. Model-Driven DTs.

2.4 Digital Twins; Applications and Research

In this section, the main applications of DTs are described. Note that the majority of them are not mature yet. Therefore, the applications described below reflect the potential of the DTs and not commercial products.

DTs are utilized by the **Healthcare sector** to optimize the patient care, support medical decisions, reduce the treatment's cost, improve the performance of medical operations and provide a personalized healthcare experience. They provide a virtualized healthcare experience which optimizes patient's care, treatment cost and performance.

DTs are utilized by the **Manufacturing sector** to perform simulations over virtual machines, develop evaluation models of product quality, improve and speed up production processes and operations, reduce production and maintenance costs, and identify failures on machines and their physical components.

DTs are utilized by **Automotive sector** to execute model-based simulations over virtual replicas of vehicles before their production, execute model-based simulations within a

virtual road environment and near real-life conditions, optimize the production lifecycle, enable autonomous capabilities of vehicles and evaluate their liability.

2.5 Differences between DT and CDT

CDT is an instance of DT that refers to humans (Consumers) within an energy market. Therefore, CDT has common characteristics inherited from the DTs. The fact that a CDT represents a human entity results to some differences which are mentioned below⁷.

- a. DTs support synchronized bidirectional transfer of data between the physical and the virtual entity. In the case of CDT, the transmission between the virtual environment of the CDT and the Consumer is mostly asynchronous, as Consumers cannot be 24/7/365 connected and synchronized with their virtual replica.
- b. CDT models human behaviour that involves subjectivity. DTs model behaviours of non-viable entities which subject to mathematical and physical laws.
- c. CDT produces less accurate outputs than DTs. In fact, the output of the CDT is probabilistic, as the Consumers' behaviour is hard to be described and modelled deterministically.
- d. CDT deals with the liquid social environment of Consumers that influence their behaviours. DTs deals with a strictly defined environment of non-viable entities.

2.6 Consumer Digital Twin; 3-tier Architecture

The CDT implements a **3-tier architecture** that consists of the tree (3) layers; Process Layer, Communication Layer and Aggregation Layer.

The **Process Layer** of CDT processes the collected raw data and transforms them into a useful and valuable form. It performs statistical methods and AI/ML techniques over data to make forecasts, predictions, discover patterns, classify data and support informed decisions. The layer's output is displayed on the front-end of CDT.

⁷ S. Wei, Is Human Digital Twin Possible? Computer Methods and Programs in Biomedicine Update, Elsevier, 2021

The **Communication Layer** is the link layer between tiers that supports the communication between CDT and data resources. It provides the appropriate bandwidth, latency, reliability and security to data transmission.

The **Aggregation Layer** acquires raw data from every available inventory resource. CDT acquires data from physical resources, such as IoT devices, smart meters and wearable devices, and digital resources, such as API services.

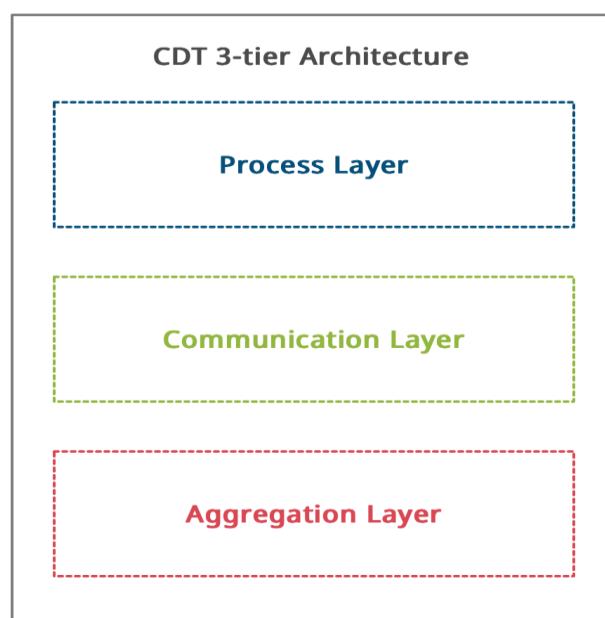


Figure 1 The 3-tier Architecture of CDT

3.CDT Platform, Technologies, Data Aggregation, Standards and Interfaces

3.1 CRISP - DM Methodology

The Cross-Industry Standard Process model for Data Mining, known as **CRISP-DM Methodology**, is utilized in the development process of Data-Driven DTs ^{8,9}. CRISP-DM Methodology provides a framework which can be universally adopted on DT projects regardless of the sector applied and the technologies used. It comprises six (6) sub-sequential steps or phases; (a): Business Understanding, (b): Data Understanding, (c): Data Preparation, (d): Modelling, (e): Evaluation and (f): Deployment as presented on Fig.3.X. In each step, specific tasks are executed which produce process instances as outputs. CRISP-DM is a flexible method, as it allows iterations between the steps, i.e., moving forwards and backwards, to improve the output of a single phase and optimize the process.

⁸ C. Schröder et.al, A Systematic Literature Review on Applying CRISP-DM Process Model, Procedia Computer Science, Elsevier, 2021

⁹ R. Wirth and J. Hipp, CRISP-DM: Towards a Standard Process Model for Data Mining, 2000

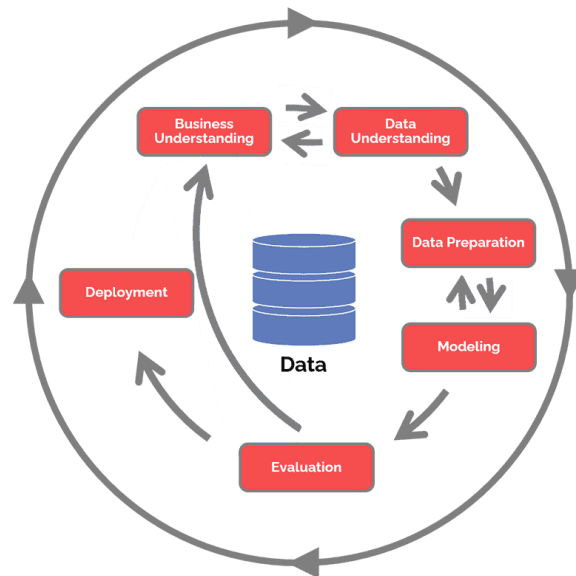


Figure 2. The steps of CRISP-DM Methodology

The development process of CDT has been based on the CRISP-DM Methodology. In the following sections, each step is described.

3.1.1 CDT: Business Understanding

The initial step of CRISP-DM methodology requires the definition of the DT's business objective and the success criteria. The model of a Consumer within an energy market is the business objective of the CDT. In addition, the necessary data resources are listed and any assumptions or constraints are specified.

3.1.2 CDT: Data Understanding

The second step of CRISP-DM methodology includes a verification process that the data resources are accessible and available. Then, tasks related to data exploration, such as data visualization and statistical analysis of data, are performed. In that way, CDT gains insights about key characteristics, attributes and surface features of Consumer. In addition, tasks related to verification of data quality are performed such as inspection for outliers and null values.

3.1.3 CDT: Data Preparation

The third step of CRISP-DM methodology includes the transformation of raw data into valuable datasets. Therefore, data are prepared to be applied on AI/ML algorithms and statistical methods. In addition, tasks related to data dimension reduction, cleansing from null values / outliers and isolation of key characteristics and features are performed.

3.1.4 CDT: Modelling

The fourth step of CRISP-DM methodology performs AI/ML algorithms and statistical models over the datasets produced at the previous step. These can be regression models, classification techniques, time series models, multicriteria decision analysis models and statistical models. The models employed by CDT are described in detail in Section 4.

3.1.5 CDT: Evaluation

The fifth step of CRISP-DM includes the evaluation of the DT to ensure that it meets the business objective. In addition, model validation tasks are performed to ensure that the core functionalities of CDT are fulfilled and the model outputs are evaluated to ensure that they meet restrictions and constraints, as defined in the first step.

3.1.6 CDT: Deployment

The final step of CRISP-DM includes the deployment of the DT. In the case of the CDT, it is deployed on the Cloud. After the deployment tasks are completed, the end-users can access the CDT. Note that the deployment of CDT does not finalize the process of CRISP-DM. Maintenance tasks and updates are required on a regular basis.

3.2 CDT Platform Technologies

3.2.1 Functional and non-Functional System Requirements

The functional requirements of the CDT specify what the CDT should do. It describes the behaviour of the CDT when certain conditions are met.

The non-functional requirements of the CDT specify the series of actions that have to be performed to achieve functional requirements or how the CDT works.

Functional Requirement 1: Authentication of the Consumer

Non-Functional Requirement 1: Implementation of Keycloak

Functional Requirement 2: Visualization of data analytics and information

Non-Functional Requirement 1: Collect and Store data utilizing the iSCAN DB

Non-Functional Requirement 2: Implementation of front-end

Functional Requirement 3: Profile of the Consumer

Non-Functional Requirement 1: Collect and Store data utilizing the iSCAN DB

Non-Functional Requirement 2: Employ AI/ML algorithms and techniques

Functional Requirement 4: Make Forecasts and Predictions

Non-Functional Requirement 1: Collect and Store data utilizing the iSCAN DB

Non-Functional Requirement 2: Employ time-series / regression algorithms

Functional Requirement 5: Support Consumer's 'Preferences'

Non-Functional Requirement 1: Implementation of the 'Preferences' front-end

Non-Functional Requirement 2: Collect and Store data utilizing the iSCAN DB

Non-Functional Requirement 3: Employ MCDM algorithms and techniques

3.2.1 CDT as a Web Application

A typical application is developed either as a desktop application or a web-based application. Nowadays, web-based applications tend to be a trend, as they have numerous advantages over desktop applications. Therefore, CDT has been developed as a web-based application. The main advantages of web-based applications for the end-users are portability, universal accessibility and cross-platform availability that make these applications run regardless of the user's location, the type of end-device and the operating system. In addition, web-based applications do not require installation and manual updates from the end-users that makes them even more attractive. Web-based applications have advantages for the development team, too. These applications are highly scalable, their deployment process is fast and easy, their maintenance is centrally performed and provide a better security status.

3.2.2 Front-end side

The CDT has been developed as a web-based application; thus, its front-end employs the core web technologies of HTML, CSS and JavaScript. To extend the functionality of the front-end, improve the performance of the front-end and provide a higher customization level over UI components, CDT utilizes additionally external open-source libraries and frameworks.

The front-end of the CDT provides an interactive GUI interface to the end-user. It visualizes information in a user-friendly and valuable form for the end-user, as well as provides an end-point to interact with the functionalities of the CDT, such as the Preferences. The interaction is supported by various UI components such as sliders, cards, tables and charts. Moreover, the front-end of the CDT is fully responsive, which means it can be adjusted dynamically to the screen size or the orientation of any device, such as desktops, smartphones and tablets.

The front-end of the CDT complies with the TwinERGY brand book guidelines. A specific CSS file has been implemented that contains the project's brand visual identity, the font-families, the colour-palette and icons. In addition, the front-end of the CDT complies with the W3 validation guidelines. It validates the quality of the HTML/CSS markup and ensures that the content of the CDT is accessible for people with disabilities.

As referred above, the CDT employs external open-source CSS and JavaScript, as described below.

CSS Libraries

CDT integrates open-source external CSS libraries to provide a user-friendly and responsive front-end to the end-user, customize the style of the UI components, comply with TwinERGY's branding guideline and achieve a qualitative aesthetic effect.

- The “**bootstrap-css**” library ¹⁰ provides a framework that simplifies the development process of responsive applications, and provides a series of modern

¹⁰ Library and documentation are available on <https://getbootstrap.com>.

UI components that can be easily customized. The UI components include jumbotrons, carousels, breadcrumbs, navbars, alerts, cards, tooltips and modals. As for the customization, there are plenty of “classes” and “properties” that can be applied to UI components to achieve the desired visual result, such as outlined buttons applied to UI buttons and stripped tables applied to UI tables.

- The “**ion-slider-css**” library ¹¹ provides customizable UI slider components. In case of CDT, they are used by the end-users to insert the importance level of each Preference.

JavaScript Libraries

CDT integrates open-source external JavaScript libraries to improve the performance in terms of computational resources and time, as well as, extend the functionalities of the CDT in the front-end, such as dynamic customization of UI components. In addition, it utilizes AJAX calls to support asynchronous HTTP requests and responses.

- The “**jQuery**” library ¹² provides a set of faster and codeless methods for HTML/DOM manipulation.
- The “**chart-js**” library ¹³ provides a collection of pre-built chart components, such as line charts, bar charts, radar charts, scatter plots. That library extends the core chart functionalities of HTML canvas tag.

3.2.2 Back-end side

The back-end side refers to the server side tasks which are performed to achieve the functionalities of the CDT. At the server side, data is collected and processed through scripts to provide a dynamic aspect of the application. The collection of data is performed through transactions with the iSCAN Database system, transactions with the CDT's dedicated Database or third-party open APIs.

¹¹ Library and documentation are available on <https://ionicframework.com/docs/api/slides>.

¹² Library and documentation are available on <https://jquery.com>.

¹³ Library and documentation are available on <https://www.chartjs.org>.

The back-end scripts of the CDT are written either in PHP or Python. Each technology is employed to perform specific tasks. PHP is utilized for transactions with the Databases and the third-party open APIs, whereas, Python focuses on data analysis and processing. Python is employed to perform AI/ML algorithms, MCDM models and statistical analysis of the available data and produce outputs related to the core functionalities of the CDT, such as Consumer's Clustering, Forecasts on Consumer's Energy Demand and Production, Consumer's Preferences Analysis, Informed Decision Support and Thermal Flexibility Analysis.

Python Libraries

CDT employs Python frameworks, packages and libraries to reduce the time and simplify the development process.

Pandas is an external open-source Python library ¹⁴ which is used to work easily with relational or labelled data. It provides built-in methods to perform tasks related to data selection, data merge, data cleansing, data fill, data normalization, data inspection, and statistical analysis.

NumPy is an external open-source Python library ¹⁵ which is used to work easily with calculations over matrices and other complex mathematical operations (e.g. Fourier Transform).

Scikit-learn is an external open-source Python library ¹⁶ which is used to execute AI/ML algorithms and statistical models. It contains a qualitative code collection that provides tools for regression analysis, classification, clustering and dimensionality reduction.

¹⁴ Library and documentation are available on <https://pandas.pydata.org>

¹⁵ Library and documentation are available on <https://numpy.org>

¹⁶ Library and documentation are available on <https://scikit-learn.org/stable/>.

3.3 Authentication through Keycloak

3.3.1 Introduction

Keycloak ¹⁷ is an open-source software product that allows Single Sign-On (SSO) with Identity and Access Management aimed at web applications and services. In TwinERGY, Keycloak has been selected as the SSO provider for accessing modules of the TwinERGY ecosystem. Keycloak supports both OpenID Connect¹⁸ (an extension to OAuth 2.0¹⁹) and SAML 2.0²⁰. When securing clients and services the first thing to be decided is which of the two is going to be used. It can be also configured that some resources are going to be secured with OpenID Connect and others with SAML. Moreover, Keycloak can be easily configured to work together with well-known external services as Identity Providers. Currently, Keycloak supports Google, Facebook, Twitter, GitHub, LinkedIn, Microsoft, Instagram, Bitbucket, GitLab, PayPal, Stack Overflow, etc. Such a capability often appears useful as users are able to use their existing digital identities from these services. Figure below explains the identity flow in case an external Identity Provider is used. Identity flow remains the same also in case when Keycloak internally undertakes the role of Identity Provider.

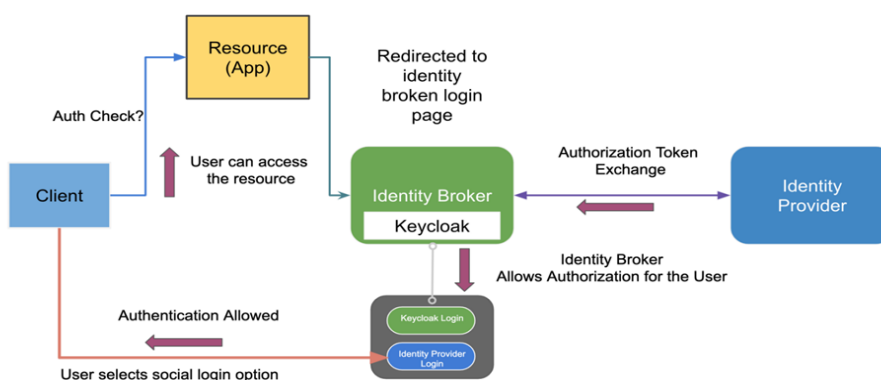


Figure 3. Keycloak identity flow explained with or without an external Identity Provider

¹⁷ <https://www.keycloak.org/>

¹⁸ http://openid.net/specs/openid-connect-core-1_0.html

¹⁹ <https://oauth.net/2/>

²⁰ <https://wiki.oasis-open.org/security/FrontPage>

Requests raised to TwinERGY modules should be redirected to the login page of the TwinERGY authentication provider (Keycloak). The user enters and submits his credentials and if everything is OK, he is redirected to the URL of his initial request. Additionally, a new user can create his identity and profile if he is not registered yet.

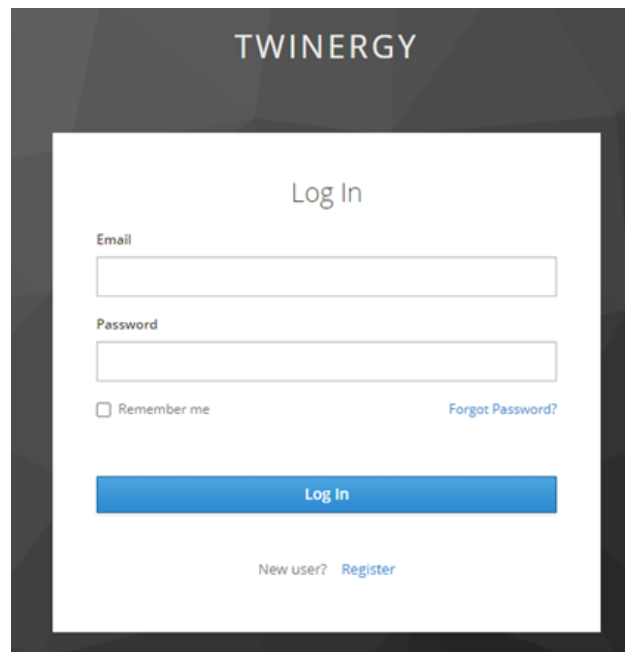


Figure 4. Example SSO login page with Keycloak

3.3.2 Keycloak configuration for TwinERGY

For convenience and regarding TwinERGY CDT module, Keycloak has been deployed and configured within a stage server in University of Patras for integration testing purposes. As the development of the CDT module gets finalized current configuration will be migrated to its final deployment.

To achieve the SSO functionality, a new realm (Figure 5), and the respective clients (Figure 6), one for each application, have been created for that purpose. The detail of the configuration is beyond the scope of this document.

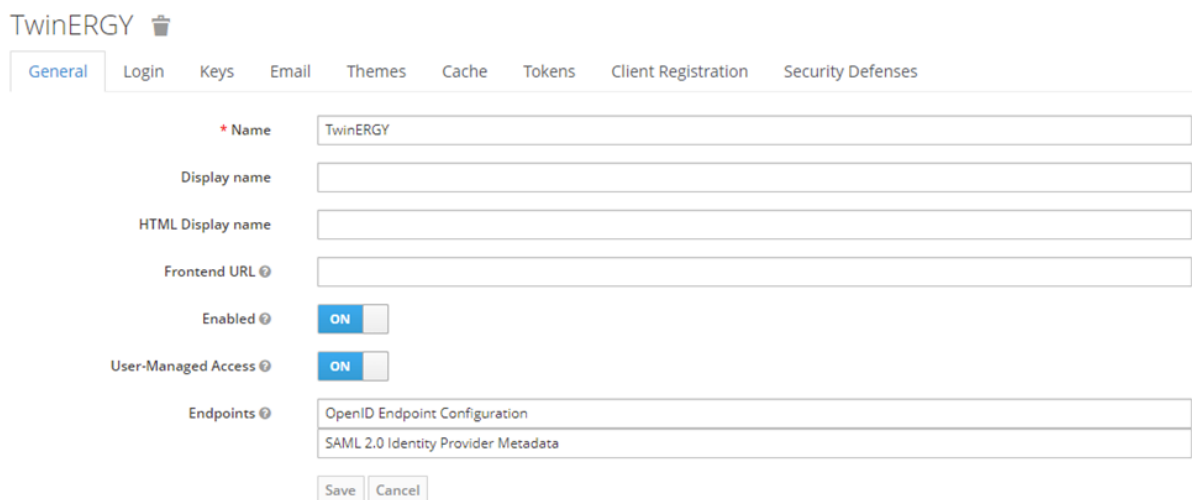


Figure 5. TwinERGY Realm

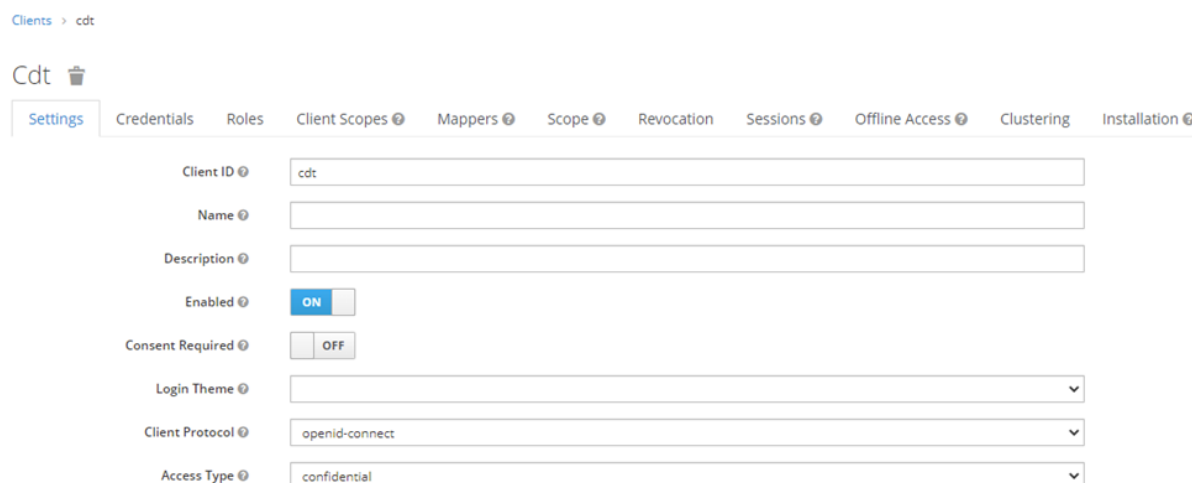



Figure 6. TwinERGY client

Any specific attributes that need to be defined can be created and assigned to users using the Mappers section of the above figure. By default, Keycloak supports a wide variety of user attributes, but they can be further extended as shown in Figure 7, Figure 8 and Figure 9, where a user-defined attribute named: **TwinERGY Attr 1** has been defined for integration testing purposes.

Cdt 


Settings Credentials Roles Client Scopes **Mappers** Scope Revocation Sessions Offline Access Clustering Installation


Name	Category	Type	Priority Order	Actions
email	Token mapper	User Property	0	Edit Delete
town	Token mapper	User Attribute	0	Edit Delete
TwinERGY Attr 1	Token mapper	User Attribute	0	Edit Delete
middle name	Token mapper	User Attribute	0	Edit Delete
profile	Token mapper	User Attribute	0	Edit Delete
address	Token mapper	User Address	0	Edit Delete
birthdate	Token mapper	User Attribute	0	Edit Delete
family name	Token mapper	User Property	0	Edit Delete
phone number verified	Token mapper	User Attribute	0	Edit Delete
given name	Token mapper	User Property	0	Edit Delete
allowed web origins	Token mapper	Allowed Web Origins	0	Edit Delete
upn	Token mapper	User Property	0	Edit Delete


Figure 7. Sample list of user attributes


Clients > cdt > Mappers > TwinERGY Attr 1


TwinERGY Attr 1 


Protocol  openid-connect


ID  fcd0c346-846e-4058-ad7c-d4aa8565cfc5


Name  TwinERGY Attr 1


Mapper Type  User Attribute


User Attribute  cdtAttr1


Token Claim Name  cdt_attr_1


Claim JSON Type  String

Add to ID token  ON

Add to access token  ON

Add to userinfo  ON

Multivalued  OFF

Aggregate attribute values  OFF

Save Cancel

Figure 8. Customization of user-defined attribute for TwinERGY scope

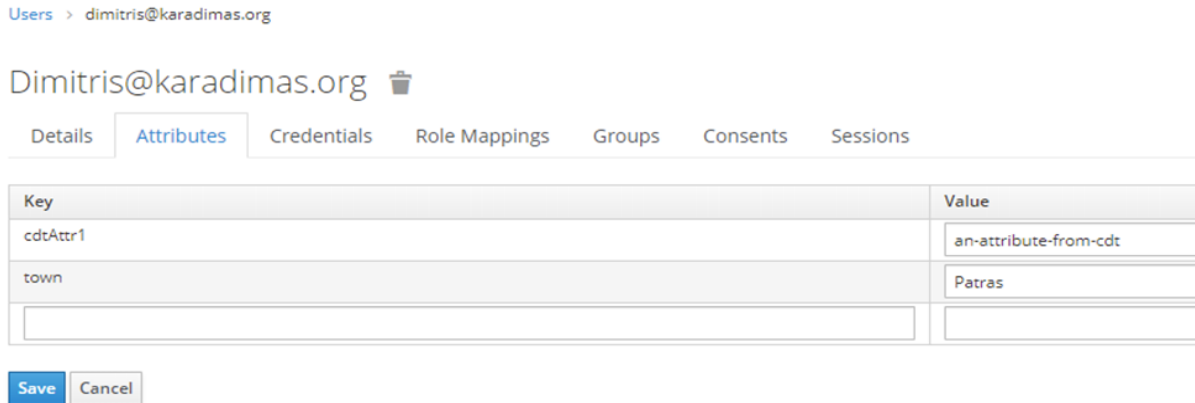


Figure 9. Assignment of a user-defined attribute for a user

Access to the TwinERGY modules and interfaces is provided following a role-based permission scheme. Specific user roles have been created on the realm's configuration page (Figure 10). These Keycloak roles have been mapped to roles defined in the respective applications (Figure 11).



Figure 10. Keycloak user roles

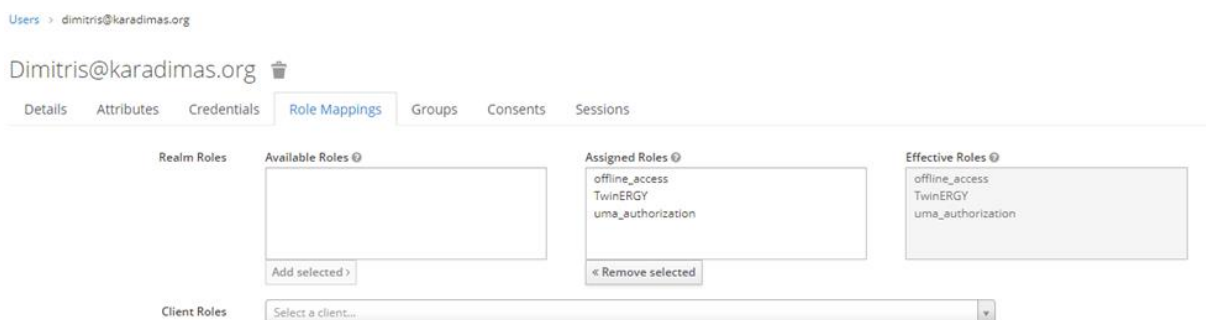


Figure 11. Role's mappings

3.3.3 Configuration for TwinERGY CDT module

To secure clients and services, we are also going to need an adapter or library for the protocol we have selected. Keycloak comes with its own adapters for selected platforms, but it is also possible to use generic OpenID Connect Relying Party and SAML Service Provider libraries.

Specifically, for the TwinERGY CDT module the *mod_auth_openidc* library for Apache HTTP Server has been selected and deployed, as it is a certified authentication and authorization module for the Apache 2.x HTTP server that implements the OpenID Connect Relying Party functionality.

This library enables an Apache 2.x web server to operate as an OpenID Connect Relying Party (RP) towards an OpenID Connect Provider (OP). It relays end user authentication to a Provider and receives user identity information from that Provider. It then passes on that identity information (a.k.a. claims) to applications protected by the Apache web server and establishes an authentication session for the identified user.

The protected content, applications and services can be hosted by the Apache server itself or served from the origin server(s) residing behind it by configuring Apache as a Reverse Proxy in front of those servers. The latter allows for adding OpenID Connect based authentication to existing applications/services/SPAs without modifying those applications, possibly migrating them away from legacy authentication mechanisms to standards-based OpenID Connect Single Sign On (SSO).

Deployment of the *mod_auth_openidc* library for Apache HTTP Server within the stage server (CentOS 7) has been realized following the next steps.

Setup Apache2 SSL, if not already done:

```
yum install httpd mod_ssl  
yum install curl hiredis jansson
```

Install the mod_auth_openidc Apache2 module:

```
#rpm -ivh
https://github.com/pingidentity/mod_auth_openidc/releases/download/v1.8.2/mod_auth_openidc-1.8.2-1.el6.x86_64.rpm

rpm -ivh
http://mirror.centos.org/centos/7/updates/x86_64/Packages/mod_auth_openidc-1.8.8-9.el7_9.x86_64.rpm
```

Verify installation of mod_auth_openidc Apache2 module:

```
ls -l /usr/lib64/httpd/modules/mod_auth_openidc.so
```

Configuration of an Apache2 Virtual Host for TwinERGY module:

```
LoadModule auth_openidc_module modules/mod_auth_openidc.so

<VirtualHost *:44443>
    ServerName dimitrisvdk
    ServerAdmin webmaster@localhost
    DocumentRoot /var/www/html/dev/keycloak-demo/

    #this is required by mod_auth_openidc
    OIDCCryptoPassphrase a-random-secret-used-by-apache-oidc-and-balancer

    OIDCProviderMetadataURL https://auth.dimitrisvdk.io/auth/realms/TwinERGY/.well-known/openid-configuration

    OIDCClientID cdt
    OIDCClientSecret 3d760118-5811-4cfa-92d2-8c041f653fe6
    OIDCRedirectURI http://dimitrisvdk:44443/protected/redirect.php

    # maps the preferred_username claim to the REMOTE_USER environment variable
    OIDCRemoteUserClaim preferred_username

    #OIDCPassClaimsAs environment
    #OIDCClaimPrefix USERINFO_
    OIDCPassIDTokenAs payload

    <Location /protected/>
        AuthType openid-connect
        Require valid-user
    </Location>

</VirtualHost>
```

Fig. 3.11 Apache2 Virtual Host configuration.

3.3.4 Realisation of a protected TwinERGY CDT module

By default, the module sets the `REMOTE_USER` variable to the `id_token [sub]` claim, concatenated with the OP's Issuer identifier. Other `id_token` claims are passed in HTTP headers and/or environment variables together with those (optionally) obtained from the

UserInfo endpoint. The provided HTTP headers and environment variables can be consumed by applications protected by the Apache server.

Figures below present an example instance from HTTP request and response headers of a protected TwinERGY resource under Keycloak.

```

Request Headers View source
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9
Accept-Encoding: gzip, deflate
Accept-Language: en-US,en;q=0.9
Cache-Control: no-cache
Connection: keep-alive
Cookie: PHPSESSID=6042c4b8a4641e54f49e24ba77ec0f77; mod_auth_openidc_session=0a0fb0f0-245a-4989-9802-1bb1aba2dea3
DNT: 1
Host: dimitris.dk:44443
Pragma: no-cache
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/99.0.4844.82 Safari/537.36
    
```

Fig. 3.12 HTTP request headers of a Keycloak protected TwinERGY resource.

```

Response Headers View source
Connection: Keep-Alive
Content-Length: 478
Content-Type: text/html; charset=iso-8859-1
Date: Tue, 22 Mar 2022 18:01:09 GMT
Keep-Alive: timeout=5, max=100
Location: https://auth.dimitris.dk/auth/realms/TwinERGY/protocol/openid-connect/auth?response_type=code&scope=openid&client_id=cdt&state=WS6rsaCTOKg17D7_yAU4FwkZ4kg&redirect_uri=http%3A%2F%2Fdimitris.dk%3A4444%3Fprotected%2Fredirect.php&nonce=XS3P-k952gwZMnpXx9jGg_Bk002XyzTceZ7YpZ26Rrs
Server: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips
Set-Cookie: mod_auth_openidc_state_WS6rsaCTOKg17D7_yAU4FwkZ4kg=Z2gCG2jYirJj0NqkKCPLsgtsBw_YILRL7c-UQy4k_H_cIX1Gf0nbX2pFDWn0jfkpvku5deTFaALTSUYX0iLpHE-DcwiIABeJ88n5sa3pw6Wzd-G8Em2QzLNiis_QKU9rZgrIIISefiNytBwCnvOm5Pun66Rb-0cXS46k6FD6fKnk0c5I9LpTnV7B0sguj27Y4x600WSD9oJrmYnKk8edIOCD7F7NDIimIT19-0ZFypqUcNu04aPoCFPu1OnZsUxooYGTqBt6CRD4f1kt1oxwPEXqW0NQung7z5xaXGbzL2c-cgc0cJ8CF1qK2J02_4FK27V-_1zYFX8-nPrpVQ;Path=/;HttpOnly
    
```

Fig. 3.13 HTTP response headers of a Keycloak protected TwinERGY resource.

```

Request initiator chain
http://dimitris.dk:4443/protected/index.php
  https://auth.dimitris.dk/auth/realms/TwinERGY/protocol/openid-connect/auth?response_type=code&scope=openid&client_id=cdt&state=WS6rsaCTOKg17D7_yAU4FwkZ4kg&redirect_uri=http%3A%2F%2Fdimitris.dk%3A4444%3Fprotected%2Fredirect.php?state=WS6rsaCTOKg17D7_yAU4FwkZ4kg&session_state=ec5f9af3-4589-4e5a-9e32-9e33120533ad&code=b5cadad1-8d63-45cc-833d-62441431aad
    https://auth.dimitris.dk:4443/protected/redirect.php
      http://dimitris.dk:4443/protected/index.php
    
```

Fig. 3.14 HTTP request chain of a Keycloak protected TwinERGY resource.

Request Cookies <input type="checkbox"/> show filtered out request cookies					
Name	Value	Domain	Path	Expires ..	Size
PHPSESSID	6042c4b8a4641e54f49e24ba77ec0f77	dimitris...	/	Session	41
mod_auth_openidc_session	0a0fb0f0-245a-4989-9802-1bb1aba2dea3	dimitris...	/	Session	60
Response Cookies					
Name	Value	Domain	Path	Expires ..	Size
mod_auth_openidc_state_WS6rsaCTOK...	Z2gCG2jYirJj0NqkKCPLsgtsBw_YILRL7c-UQy4...	dimitris...	/	Session	409

Figure 12. Request cookies of a Keycloak protected TwinERGY resource

```

{
  "realm": "TwinERGY",
  "public_key": "MIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEAI3TOIbhmPimGK1vORpAxs8X+jeoPznrpekgsu2X0Fh+Z11Ku119RUj1Hgapt3qEQqGU1ngh17utdP/pVZvFhC8jcyDjHPS1tBZC5iRoAkS8sCyv7kGB1GwDClkvFxoVhniVfihVrmPen8XoduskZH7uF7XsJtscfrK3w053Xtvi76X0tzqXyc090iLkUe40QXlv6SIFvL17u0X3jgTkur31/2wUkFKJ0g4Xv70jx43/71KTRjkd4fwYUzisc/vkQj3CtnSELXBmshXKGA2XaxRgSuxl1fPEj8dEgAPXBFfpv9qWQXsCiwHieQ6Vfc/u6pBf0I4bzHvVa8B0vWIDAQAB",
  "token-service": "https://          /auth/realms/TwinERGY/protocol/openid-connect",
  "account-service": "https://          /auth/realms/TwinERGY/account",
  "tokens-not-before": 1644336956
}

```

Figure 13. Example JWT response

3.4 Interoperability (iSCAN)

Interoperability between modules and sub-systems within the TwinERGY ecosystem is supported by iSCAN. iSCAN is a cloud-based data management and interrogation platform that can import time-series data from multiple sources to a single platform, e.g., from BMS, smart meters, IoT sensors, etc. iSCAN can interrogate the data manually and supports data analysis. Furthermore, iSCAN can identify how the building is performing in real-time with respect to energy use, indoor environmental quality and running costs. Insights and analysis can then be performed on the data to identify quick wins such as faults or easy to implement energy saving measures and Machine Learning and AI can be used to optimize performance on a day-to-day basis. Time series data in iSCAN can be exported to ICL-DT ecosystem to be exploited by all technologies within the ecosystem. When this is carried out, those DTs are now a hybrid Digital Twin based on both physics data and metered data to create a higher level of accuracy and an ability to carry out more detailed analysis with respect to optimization and intervention planning.

3.5 Graphical User Interfaces & Dashboard

CDT incorporates GUIs to visualize the collected data from the Consumer’s environment and acquired knowledge from the process of collected data. CDT provides dedicated GUIs with insights and analytics for Consumer’s Energy Production, Energy Consumption, Energy Storage, Comfort Level and Well-being Status. At the same time, they display knowledge such as personalized recommendations and energy forecasts. Below, is provided an example that visualizes the ‘Energy Production’ for an individual.

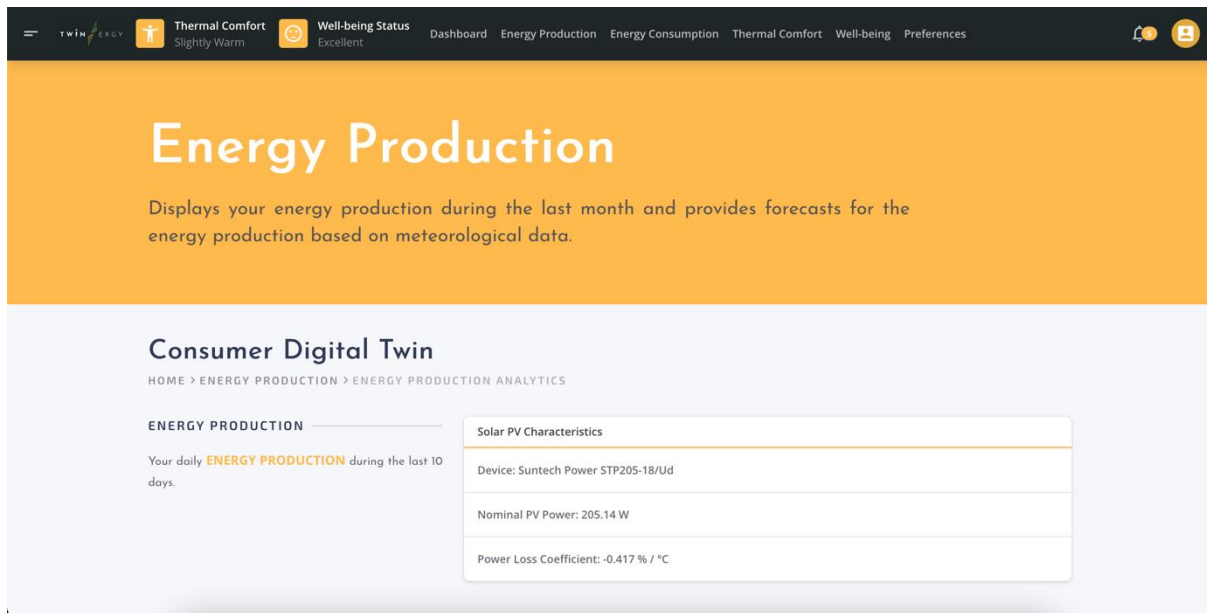


Figure 14. Information about the install Solar PV within the occupancy

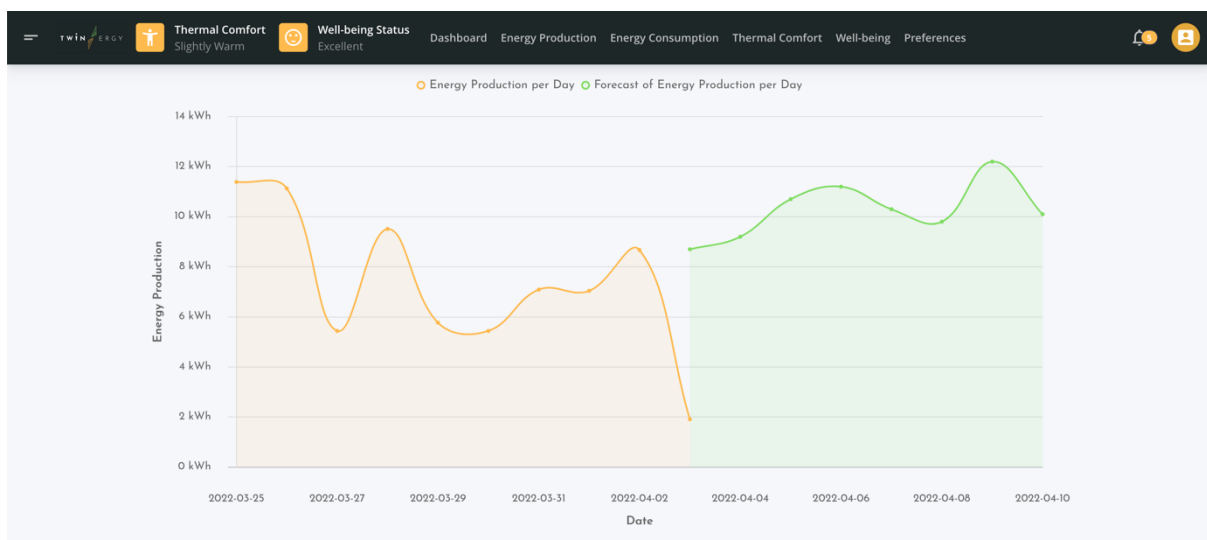


Figure 15. Aggregated daily and 7-day forecasted energy production

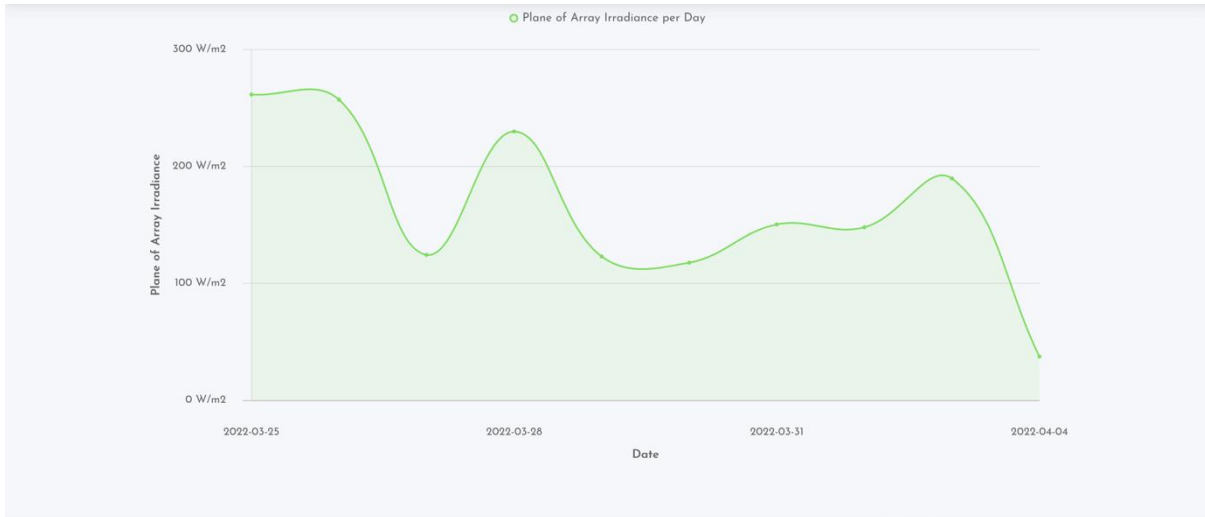


Figure 16. Environmental Conditions that affect the energy production level (1/2)



Figure 17. Environmental Conditions that affect the energy production level (2/2)

The core GUI component of the CDT is the dashboard that provides an overview of Consumers’ data, behaviors and indicators in an intuitive and user-friendly way. More specifically, it displays summarized information about the Consumer’s behavior within the energy market and provides insights about energy production, energy consumption, energy storage, thermal comfort and well-being indicators. Dashboard utilizes UI components such as cards with embedded charts to simplify the information provided to the end-user. The structure of the dashboard comprises a three-section layout, as described below, which can be modified or extended to achieve the highest engagement rate with Consumers.

In the first section, the dashboard displays information about the Consumer’s Thermal Comfort level as determined from the PMV method and the latest Activity Level as captured from the wearable device. In addition, it provides information about key factors

that affect the Thermal Comfort (Indoor Air Temperature and Indoor Relative Humidity) in the form of time-series graphs.

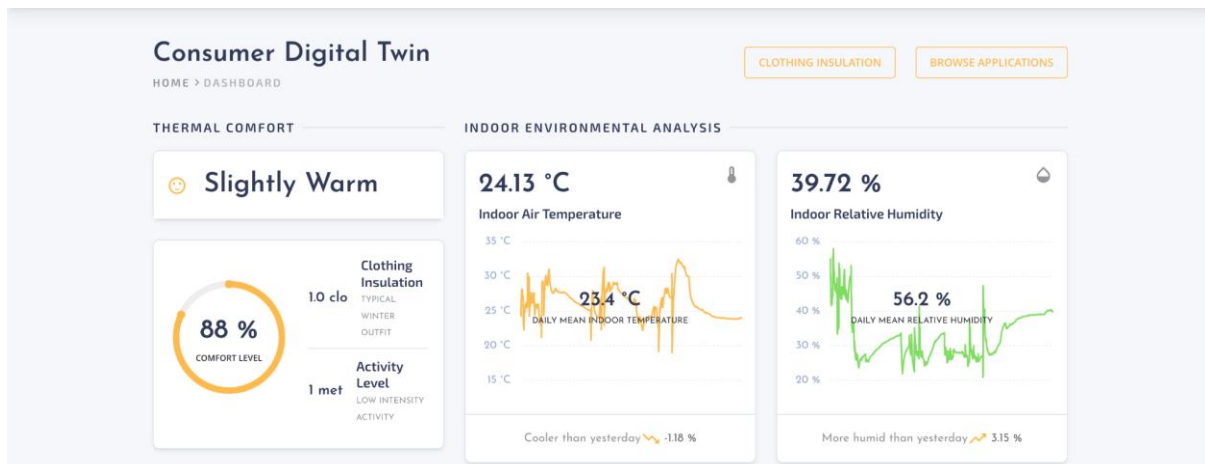


Figure 18. Dashboard of the CDT [1/3]

In the second section, it displays information about the Consumer's Well-being Status as determined from the Indoor Air Quality Model. In addition, it provides information about key factors that affect the Well-being Status (Indoor eCO₂ and TVOCs concentration) in the form of time-series graphs.

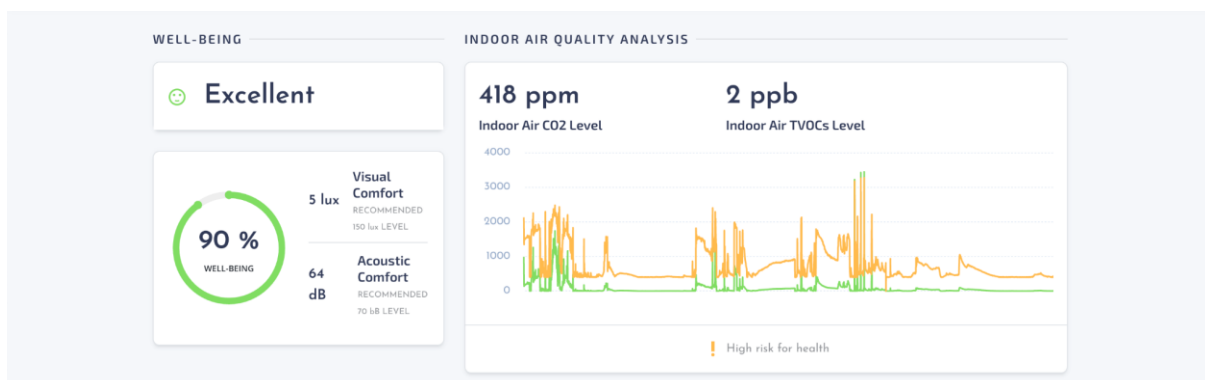


Figure 19. Dashboard of the CDT [2/3]

In the third section, it displays information related to the Consumer's Energy Production and the financial profit gained from it in the form of a line chart. In addition, it displays key environmental factors that affect the level of Energy Production, such as Cloud Coverage and Sunny Period.

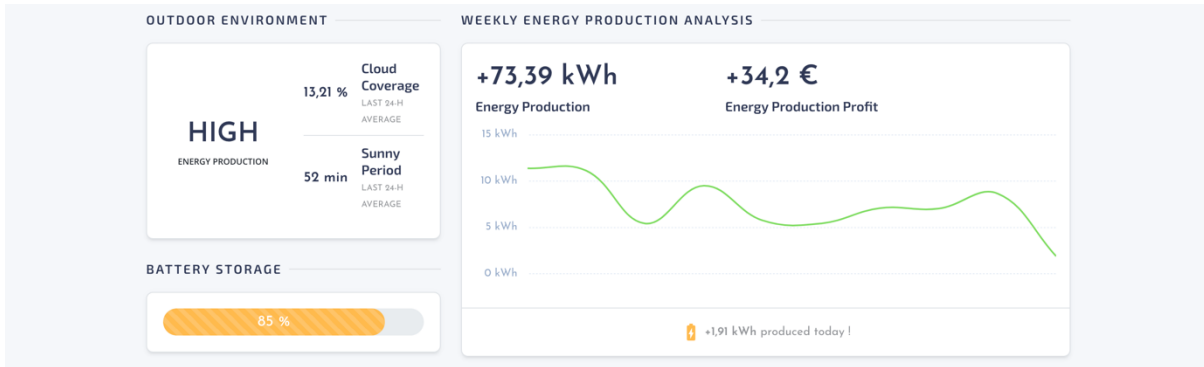


Figure 20. Dashboard of the CDT [3/3]

CDT employs another important GUI component, “Notifications” which is a drop-down list placed on the navigation menu of the application. They display recommendations to Consumers to support their decisions and information resulted as knowledge from the CDT’s algorithms. A typical example is the forecast of the energy production level, as presented on the Section 4.3, which has impact on the Consumers’ energy behaviour and decisions.



Figure 21. The notifications component of CDT

4.CDT Data Processing

4.1 Consumption Profiling

4.1.1 K-Means algorithm for Consumer profiling

Introduction to Clustering

Clustering or Cluster Analysis is an unsupervised ML technique that divides an unlabelled population or data collection into a number of groups. The groups consist of data points with similar characteristics or properties. Clustering aims to produce meaningful clusters of data points which are characterized by homogeneity within a cluster and heterogeneity between different data clusters. Clustering has been adopted by various market sectors to identify customer segments, acknowledge customer behaviours, classify customers with similar characteristics and profile customers. In the scope of the CDT, it is utilized to divide Consumers into segments based on their behavioural characteristics, features and preferences.

K-means clustering

Clustering employs various algorithms to group data. In the case of the CDT, we propose the K-means algorithm as the clustering tool. The K-means algorithm is a popular unsupervised ML technique that creates clusters from a population or a collection of data. It is widely used in clustering problems, as it is easily implemented, faster than other clustering techniques and requires fewer computations and steps. K-means divides the population or data into a set of k-groups, where k is the number of pre-defined groups. The centre of a cluster has to minimize the distance between the data points within the cluster, as compared to another cluster's centroid. The most challenging task of the K-means algorithm is the determination of the right k-value. There are two approaches to select the proper k-value; Elbow Method and Silhouette Method.

The Elbow Method is an empirical method that picks up the range of potential k-values and selects the best among them. It calculates the sum of the square of the points and calculates the average distance for each k-value. Then, a graph is displayed that seems

like an elbow. The point where the curve of the graph decreases abruptly is considered as the best k-value. The Silhouette Method calculates the average distance of points within a cluster and the average distance of the points to its next closest cluster called. Then, the Silhouette coefficient is determined for each k-value, and the best k-value is considered the “k” that maximizes the coefficient, called the best-case scenario.

Clustering and CDT

As mentioned above, the most challenging part of the clustering process is the determination of the k-value. In case of multi-dimensional data, the problem of complexity should be addressed, too. In the scope of CDT, each Consumer is described by multiple characteristics and features which increases the complexity of the clustering problem. Therefore, Principal Component Analysis (PCA) technique is employed to reduce the features that describe each Consumer. PCA reduces the dimensionality of a large data frame without a significant information loss.

Let’s examine a simplified hypothetical problem. Assume that a Consumer is described by nine (9) characteristics. In fact, CDT collects much more than nine characteristics. The collected features and characteristics are shown on Table 4.1 for three (3) Consumers.

#	F ₁	F ₂	F ₃	F ₄	F ₅	F ₆	F ₇	F ₈	F ₉
1	25	35 m ²	9 h	24 °C	40 %	500 kWh	0 kWh	50 \$/m	120 \$/m
2	29	55 m ²	7 h	24 °C	45 %	483 kWh	0 kWh	60 \$/m	200 \$/m
3	40	80 m ²	1 h	27 °C	40 %	789 kWh	200 kWh	80 \$/m	200 \$/m

Table 3. A typical dataframe of Consumers and their collected features and characteristics

Assume that these characteristics refer to Consumer’s age, the domicile’s area, the Consumer’s flexibility in hours, the ideal indoor air temperature, the ideal indoor relative humidity, the mean monthly energy consumption and energy production, the lower and upper boundary for energy spendings per month. To produce meaningful clusters of Consumers, the dimensions of the data frame have to be reduced into a single dimension. It can be achieved by performing PCA technique to the initial dataframe (Table 3) which produce a reduced data frame, as shown on Table 4.

#	F'
1	1.462
2	1.748
3	1.934

Table 4. Reduced dataframe as produced by PCA method

The dimensions of the data frame reduced from nine (9) to one (1). The complexity has been reduced and the implementation of the clustering turns into a simple process. Let's note two points. First, in a real case scenario of the CDT, the Consumer's features are expected to be more than nine (9), so a dimension's reduction technique is necessary to produce proper clusters of Consumers. Second, the K-means and clustering algorithms produces ONLY clusters. **The meaning of these clusters is a problem that has to be addressed by humans. That's a project goal.**

4.1.2 Time-series model to forecast energy demand

The Demand Forecasting Methods can be classified into two (2) categories; Qualitative Method and Quantitative Methods.

The **Qualitative Methods** do not rely on rigorous mathematical methods or models, but on judgements, opinions, intuition and personal experiences. Thus, these methods retain the risk of inaccurate forecasts due to subjectivity. CDT does not utilize Qualitative Methods.

The **Quantitative Methods** rely on mathematical methods or models that are insensitive to subjectivity. There are two (2) quantitative approaches for forecasts; Time-Series and Associative Models. Time-Series Models utilizes data from the past to predict the future. It analyzes past behaviours to discover patterns and upon these patterns underlie the forecast. Associative Models utilizes parameters from the environment, related to the forecasted variable. The forecasts are based on these associations exclusively.

In case of CDT, a Quantitative Method is employed. More specifically, we use a Time-Series Regression Model with seasonality trends. Seasonality is the key of our research,

as we assume that the energy demand follows a seasonal trend within a year. More specifically, we assume that outdoor environmental conditions influence the indoor environmental conditions, thus, in seasons with extreme outdoor temperatures, Consumers' energy demands are increased to reach their optimal thermal comfort level. To maintain a thermally comfortable environment, Consumers use HVAC systems, either as cooling or heating systems depending on the season. According to studies, HVAC systems can contribute to the total energy consumption of a domicile up to 70%. So, we can conclude that the energy demand has a seasonal pattern.

Case Study

Let's examine a Case Study of the energy demand for a single Consumer. We assume that aggregated energy consumption data for each season is available, as presented at the Table 5. More specifically, two-year past data are available, and each year is divided into four (4) seasons, i.e., I, II, III, IV. For each season, the Total Demand [Y] is available.

The **Mean Seasonal Demand (MSD)** is the mean Energy Demand [Y] of a specific Season. For example, the MSD for Season I is the average of Demand of Period 1 and Period 2.

The **Seasonal Factor** is the ratio between the MSD of a specific Season and the Mean Demand. For example, the Seasonal Factor for Season I is calculated as the ratio; $1350 \text{ kWh} / 2212 \text{ kWh} = 0.61$.

The **Demand without Seasonality** is the ratio between the Demand [Y] of a specific Period and the Seasonal Factor.

[X]		[Y]			[Y _d]
Period	Semester	Demand	Mean Seasonal Demand	Seasonal Factor	Demand without Seasonality
1	I	1200 kWh	1350 kWh	0.61	$1200/0.61 = 1967 \text{ kWh}$
2	II	1900 kWh	1950 kWh	0.88	$1900/0.88 = 2159 \text{ kWh}$
3	III	3000 kWh	2850 kWh	1.29	$3000/1.29 = 2325 \text{ kWh}$
4	IV	2500 kWh	2700 kWh	1.22	$2500/1.22 = 2049 \text{ kWh}$
5	I	1500 kWh		0.61	$1500/0.61 = 2459 \text{ kWh}$

6	II	2000 kWh		0.88	2000/0.88 = 2272 kWh
7	III	2700 kWh		1.29	2700/1.29 = 2093 kWh
8	IV	2900 kWh		1.22	2900/1.22 = 2377 kWh
Sum					
36		17700 kWh			17701 kWh
Mean					
4.5		2212 kWh			2212 kWh

Table 5. Consumer's Energy Demand for the past two years

The **Time-Series Regression Model** describes a linear relationship between the independent variable of Period [X] and the dependent variable of Demand [Y]. For a specific time Period [x], the Demand Y_t can be calculated from the Equation 1:

$$Y_t = a + bx, \text{ where } a = \bar{y} - b\bar{x} \text{ and } b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

Equation 1. Linear Regression model

By applying the data into the Equation 1 we determine the parameters a and b. For the specific Case Study, the parameter a is equal to 498.56 and parameter b is equal to 39.64, so the model can be described from the Equation 2:

$$Y_t = 498.56 + 39.64x$$

Equation 2. The equation that describes the energy demand

The forecast of the demand for a specific season results from the substitution of the variable x with the value of season (Period). The result describes the forecast without seasonality trends. To forecast the demand with seasonality trend, the result of Equation 2 for a specific season, has to be multiplied by the Seasonality Factor of the Period.

Period	Season	Demand without seasonality [Y_t]	Demand with seasonality
9	I	2363.84 kWh	0.61 x 2363.84 = 1441.9 kWh
10	II	2397.60 kWh	0.88 x 2397.60 = 2109.8 kWh
11	III	2431.36 kWh	1.29 x 2431.36 = 3136.4 kWh

12	IV	2465.12 kWh	$1.22 \times 2465.12 = 3007.4$ kWh
		Forecast without Seasonality Factor	Forecast with Seasonality Factor

Table 6. Forecast of Energy Demand with Seasonality Trend

Forecast of Energy Demand

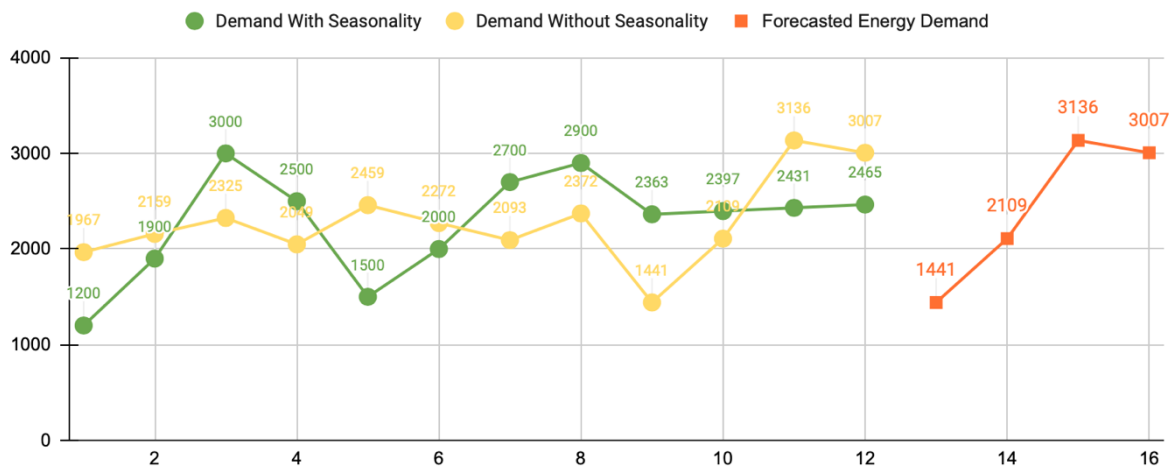


Figure 22. Forecast of Energy Demand with Seasonality Trend

4.2 Production Profiling

The data that describes the energy consumption of a Consumer is not useful itself, as many Consumers produce energy, too. In this case, Consumers are usually described with the term 'Prosumers'. The most common energy production method within a domicile is through the installation of rooftop Solar PV Panels. An accurate energy profiling of Consumers has to take into account the two aspects of a Consumer within an energy market environment; Consumer and Prosumer. Therefore, a forecast of the energy production is necessary to determine the energy balance of a Consumer and subsequently gain knowledge about their behaviour within the energy market. For example, an energy surplus can lead to higher energy flexibility and lower energy costs, whereas, an energy deficit can lead to lower thermal flexibility and higher energy costs.

CDT provides a dedicated GUI to Consumers from which they can insert the characteristics of their owned Solar PV Panels. Alternatively, they can choose the model from a predefined list of commercial Solar PVs. In that way, CDT acquires the appropriate

data to be applied into the forecast model, as proposed in the next subsection. In addition, CDT parses environmental data from an open API endpoint which are necessary to make the forecast, too.

4.2.2 Forecast the production of a Solar PV Panel

In this section, the proposed model for the forecast of energy production is described. The model requires two (2) types of inputs; environmental data and Solar PV characteristics, as shown on the Table 7.

Environmental Data	Solar PV characteristics
Ambient Temperature (T_{amb}) in units of °C	Nominal PV Power (P_{stc}) in W
Wind Speed (W_s) in units of m/s	Power Loss Coefficient (γ) in %/°C
Plane of array irradiance (G) in units of W/m ²	

Table 7. The required data for the forecast model

Forecast Energy Production of a Solar PV Panel - Step 1: Calculate PV Temperature

In the initial step, the PV Temperature is determined from the Equation 3²¹. In case the value of wind speed is not available, the PV Temperature is determined from the Equation 4²².

$$T_{PV} = T_{amb} + \omega * \frac{0.32 * G}{8.91 + 2 * \frac{W_s}{0.67}}, \text{ where } \omega = 1$$

Equation 3. Equation that describes the T_{PV} when the value of wind speed is available

$$T_{PV} = T_{amb} + k * G, \text{ where } k = 0.03 \text{ K} / \frac{\text{W}}{\text{m}^2}$$

Equation 4. Equation that describes the T_{PV} when the value of wind speed is not available

Forecast Energy Production of a Solar PV Panel - Step 2: Calculate DC Energy

In the second step, the DC energy is determined from the Equation 5.

²¹ E. Skoplaki, A. G. Boudouvis, and J. A. Palyvos, "A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting," Sol. Energy Mater. Sol. Cells, vol. 92, no. 11, pp. 1393-1402, 2008

²² R. Ross, "Interface design considerations for terrestrial solar cell modules," 1976, pp. 801-806.

$$P_{DC} = \frac{G}{1000} * P_{stc} * (1 + \gamma * (T_{PV} - 25))$$

Equation 5. Equation that describes the P_{DC}

Forecast Energy Production of a Solar PV Panel - Step 3: Calculate AC Energy

In the final step, the AC energy is determined from the Equation 6. Note that the equation requires as inputs the inverter efficiency.

$$P_{AC} = P_{DC} * n_{inverter} \text{ where } n_{inverter}: \text{ inverter efficiency}$$

Equation 6. Equation that describes the P_{AC}

4.3 Preferences; Determination of Importance

Consumers' Preferences, referred also as Preferences, are an important information source of the CDT. Preferences reflect the subjective importance between conflicting criteria related to the energy behavior of a Consumer within an energy market. These criteria describe the importance of maintaining the ideal thermal (**Thermal Comfort**) and well-being (**Well-being**) conditions within the domicile, the importance of being flexible within the energy market (**Energy Flexibility**), the importance of reducing the ecological footprint results from the energy consumption (**Eco-Friendliness**) and the importance of reducing the energy costs (**Financial Costs**). In addition, the criteria can be extended to include the importance of heavy loads in the daily routine of a Consumer. As for now, CDT includes two (2) heavy loads; the **Electric Water Heater** and the **Freezer**. Electric Water Heater reflects the importance of access to heat water during the day and Freezer reflects the importance of the freezing goods in their life. At the same time, Preferences provides a tool to Consumer to specify the ideal range for the Indoor Temperature, the desired temperature of the Electric Water Heater and Freezer, as well as, the upper boundary of the monthly energy cost.

CDT provides a dedicated GUI to Consumers, as presented on the Figure 23 and Figure 24, to insert and update their Preferences by moving a slider over pointed importance scales. They can perform actions over the Preferences in a simple and user-friendly way that acts as an incentive effect to increase the engagement rate with Consumers.

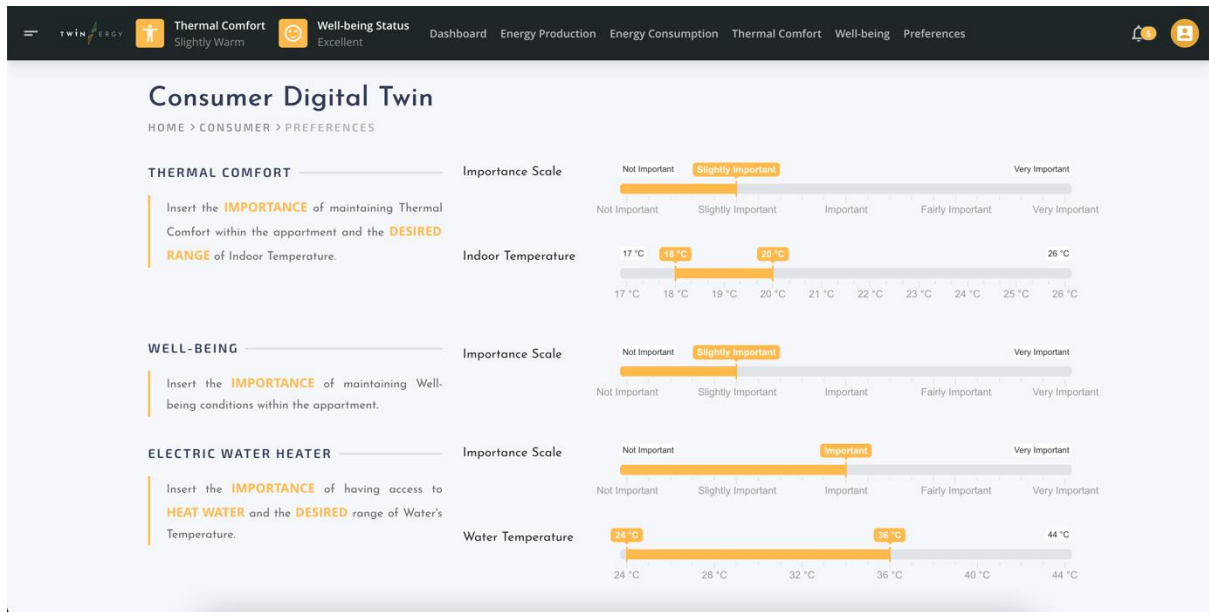


Figure 23. The GUI for Preferences (1/2)

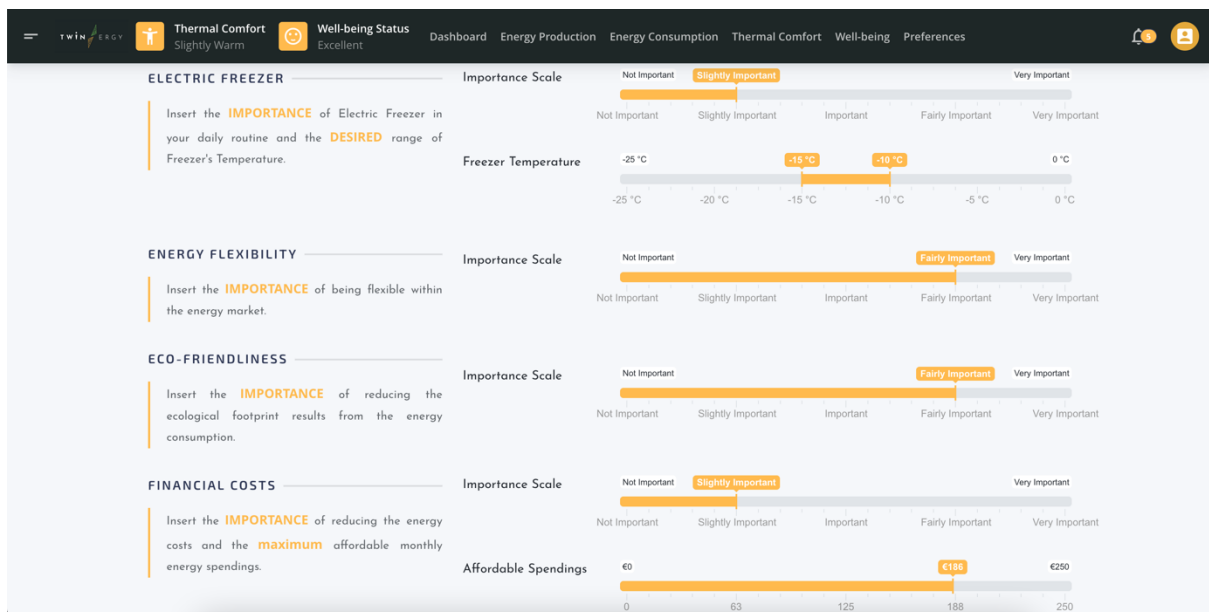


Figure 24. The GUI for Preferences (2/2)

The collected data from the Preferences are utilized either to perform MCDM methods or optimize algorithms of related modules (such as the LEM Module). So, the Preferences adds value for Consumers, as they support and optimize their decision making, empower the knowledge base for the Consumer and improves the Consumer's profiling process.

At the initialization step of the CDT, the Consumer defines the importance level of each Preference. More specifically, the Consumer inserts a value for each Preference that reflects the relative importance of the Preference in comparison with another one. In other words, a pairwise comparison takes place when the CDT is executed for the first time. It includes comparisons for all the possible combinations of Preferences. In that way, the CDT acquires useful information for the profile of the Consumer to support decisions based on the relative importance of each Preference. Then, the Consumer inserts a value for each Preference unequivocally - not in pairs -. Finally, the Consumer defines the desired range of Indoor Temperature and the maximum affordable energy cost per month. Note that Consumers can update their Preferences at any time at the lifecycle of the CDT.

Preferences
Thermal Comfort [Importance]
Indoor Temperature [Range]
Well-Being Status [Importance]
Energy Flexibility [Importance]
Eco-Friendliness [Importance]
Financial Savings [Importance]
Maximum Affordable Energy Cost [Value]

Table 8. The Preferences of CDT

4.3.1 Determine the Weight of Each Preference

The methods utilized for the determination of each Preference weight are Equal Weights (EQ) Method, Relative Importance (RI) Method, Rank Order Centroid (ROC) Method and Simos Game Method. A special method, known as Analytical Hierarchy Process (AHP) Method, is utilized to determine the weights from the pairwise comparison between Preferences.

For the pairwise comparison between Preferences, a balanced two-way 9-points scale is provided to the Consumer.

For the unequivocal determination of each Preference importance, a 5-points scale is provided to the Consumer.

Method	Scale	Comparison
Equal Weights	5-points scale	Unequivocal comparison
Relative Importance	5-points scale	Unequivocal comparison
Rank Order Centroid	5-points scale	Unequivocal comparison
Simos Game	5-points scale	Unequivocal comparison
Analytical Hierarchical Process	9-points scale	Pairwise comparison

Table 9. The point-scales used to establish the priority of criteria

Case Study

As referred above, CDT provides a 5-point scale to Consumers in order to insert the importance of each Preference. Each scale's point describes a specific importance level; (1): Not Important, (2): Slightly Important, (3): Important, (4): Fairly Important, (5): Very Important.

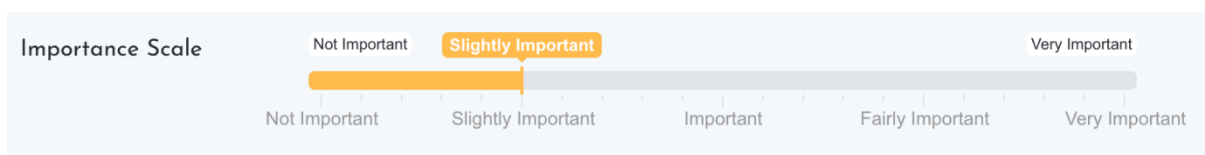


Figure 25. Importance scale for Preferences

Assume that a Consumer has inserted the importance level for the following Preferences, as shown on the Table 4.8.

Criterion	Importance Level	Importance Description
Thermal Comfort	4	Fairly Important
Well-Being	3	Important
Electric Water Heater	3	Important
Freezer	1	Not Important

Energy Flexibility	3	Important
Eco-Friendliness	1	Not Important
Financial Savings	5	Very Important

Table 10. The importance level of each Preference for the Case Study

4.3.1.1 Method 1: Equal Weights (EW)

The **Equal Weights (EW) Method** is the simplest method to determine the weights of criteria. It does not require input from the stakeholder (Consumer) about the priority of each criterion. All the criteria are considered to have the same importance which leads to equal weights for the criteria. Therefore, the EW Method can be only applied on the default state of the CDT, in which the Consumers (stakeholders) have not inserted any specific input related to the importance of each Preference (criterion).

Assume a problem of n-criteria. The weights are determined through the EW Method. The weight of each criterion (C_j) is calculated from the Equation 7 :

$$W_j(EQ) = \frac{1}{n}$$

Equation 7. Determination of each Preference weight through Equal Weights Method

In the Case Study, the Consumer (stakeholder) has inserted the importance level for five (7) Preferences (criteria). Thus, the value of n is equal to 7, and the weight of each criterion is $W_j(EQ) = 1/7$.

Preferences	Criterion (C_j)	Importance (E_j)	Weight EQ (W_j)
Thermal Comfort	C_1	4	1/7
Well-Being	C_2	3	1/7
Electric Water Heater	C_3	3	1/7
Freezer	C_4	1	1/7
Energy Flexibility	C_5	3	1/7
Eco-Friendliness	C_6	1	1/7
Financial Savings	C_7	5	1/7

Sum	1.0
-----	-----

Table 11. The weight of each Preference as determined from the EQ Method

4.3.1.2 Method 2: Relative Importance (RI)

The **Relative Importance (RI) Method** improves the determined weights from the EW Method. In contrast with the EW Method, the RI Method requires input from the stakeholder (Consumer). The stakeholder defines the importance of each criterion over a 5-points scale, as presented above. Based on the relative importance of each criterion against the other, the RI Method determines the weights.

Assume a problem of n-criteria. The weights are determined through the RI Method. The weight of each criterion (C_j) is calculated from the Equation 8 :

$$W_j(RI) = \frac{E_j}{\sum E_j}$$

Equation 8. Determination of each Preference weight through Relative Importance Method

Criterion (C_j)	Importance (E_j)	Weight EW (W_j)	Weight RI (W_j)
C_1	4	1/7	$4/E = 4/20 = 0.20$
C_2	3	1/7	$3/E = 3/20 = 0.15$
C_3	3	1/7	$3/E = 3/20 = 0.15$
C_4	1	1/7	$1/E = 1/20 = 0.05$
C_5	3	1/7	$3/E = 3/20 = 0.15$
C_6	1	1/7	$1/E = 1/20 = 0.05$
C_7	5	1/7	$5/E = 5/20 = 0.25$
Sum	20	1.0	1.0

Table 12. The weight of each Preference as determined from the RI Method

4.3.1.3 Method 3: Rank Order Centroid (ROC)

The **Rank Order Centroid (ROC) Method** establishes a weight to each criterion that minimizes the maximum error of each weight. The method identifies the centroid of all possible weights in respect to the rank order of criteria priority that is defined by the Stakeholder (Consumer). ROC Method improves even more the RI Method. The weights

of criteria defined as more important are upscaled, whilst, the weights of criteria defined as less important are downscaled.

The ROC Method has a strong drawback. It cannot be applied properly in case that two or more criteria have the same priority for the stakeholder, i.e., the same importance (E_j). In our case study, the criterion C_2 and the criterion C_3 , have the same priority for the Consumer, therefore, the ROC Method cannot be properly applied.

Assume a problem of n -criteria. The weights are determined through the RI Method. The weight of each criterion (C_j) is calculated from the Equation 9:

$$W_i = \frac{1}{n} \sum_{j=1}^n \frac{1}{j}, j = 1, 2, \dots, n$$

Equation 9. Determination of each Preference weight through Rank Order Centroid Method

Criterion (C_j)	Importance (E_j)	Weight EQ (W_j)	Weight RI (W_j)	Weight ROC (W_j)
C_1	4	1/7	0.20	$1/7 \cdot (1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7) = \mathbf{0.23}$
C_2	3	1/7	0.15	$1/7 \cdot (1/3 + 1/4 + 1/5 + 1/6 + 1/7) = \mathbf{0.16}$
C_3	3	1/7	0.15	$1/7 \cdot (1/4 + 1/5 + 1/6 + 1/7) = \mathbf{0.12}$
C_4	1	1/7	0.05	$1/7 \cdot (1/6 + 1/7) = \mathbf{0.04}$
C_5	3	1/7	0.15	$1/7 \cdot (1/5 + 1/6 + 1/7) = \mathbf{0.07}$
C_6	1	1/7	0.05	$1/7 \cdot (1/7) = \mathbf{0.01}$
C_7	5	1/7	0.25	$1/7 \cdot (1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7) = \mathbf{0.37}$
Sum		1.0	1.0	1.0

Table 13. The weight of each Preference as determined from the ROC Method

4.3.1.4 Method 4: SIMOS Method

The SIMOS Method differs from the previous methods, as it determines the weights of criteria from a five-step process - not a mathematical formula -. This method provides two main benefits. First, the criteria with equal importance for the Stakeholder have the equal weights, in contrast with the RI Method. Second, the method considers the

importance gap between criteria. Thus, intermediate importance values contribute to the final weight values.

Simos Method - Step 1

The name of each criterion (i.e., Preference) is recorded on a card. In addition, a set of blank cards are provided to the stakeholder (Consumer) as wildcards.

Simos Method - Step 2

The Stakeholder ranks the non-blank cards, i.e., the cards that represent criteria, in ASC importance order. The less important cards come first and the most important cards at the end. The blank cards that reflect the importance gap between level are placed between the non-blank cards at the empty levels. The cards that represent criteria with the same priority for the stakeholder are grouped together.

Simos Method - Step 3

A position is assigned to each criterion. Note that two or more criteria can hold a position at the same level.

Simos Method - Step 4

The non-normalized weights are determined for each rank. The non-normalized weight of each rank results as the ratio between the sum of the positions in the rank and the number of criteria in the rank.

Simos Method - Step 5

The normalized weights are determined for each rank. The normalized weight follows the standard normalization process of a sample.

Rank	#criteria	Position	Non-normalized Weights	Normalized Weights	Class Weight	Weight
CL ₁	2	1, 2	1.5	5	10	0.05
CL ₂	1	(3)	-	-	-	-

CL₃	3	3, 4, 5	4	13.7	41	0.14
CL₄	1	5	5	17.2	17	0.17
CL₅	1	6	6	20.7	21	0.21
Sum	8	29	-	-	89	0.85

Table 14. The weight of each Preference as determined from the Simos Method

The Simos Method has a strong drawback. The weights do not sum “1.0” as expected. In our Case Study, the Weights sum “0.85”. That inserts an error which can be either acceptable or non-acceptable. Therefore, a revised Simos Method has been proposed in the literature. However, in the scope of our research, the results of the non-revised method are satisfactory.

4.3.1.5 Method 5: Analytical Hierarchy Process (AHP)

The **Analytical Hierarchy Process (AHP) Method** approaches the determination of weights in a pairwise manner. The stakeholder (Consumer) inserts the priority of each criterion over another one. The decision criteria are compared in pairs. The sensitivity of importance results from the pairwise comparison across a 9-points absolute scale, as presented previously.

In case of two criteria with the same (equal) importance for the stakeholder, the value over the scale is “1”. In case a criterion is more important than the one compared with, then the stakeholder has to define the relative importance level. Each importance level is represented by a discrete value [3, 5, 7, 9] over the scale. Note that the scale is two-way. So, if the value lies on the left half of the scale, it indicates a preference of the left-side criterion over the right-one and vice versa. The intermediate values of [2, 4, 6, 8] are not used, as these values indicate neutrality.

Value	Definition	Explanation
1	Equal importance	Two preferences contribute equally to the objective

3	Moderate importance	Moderate favour of one preference over the other
5	Strong importance	Strong favour of one preference over the other
7	Very Strong importance	Very Strong favour of one preference over the other
9	Extreme importance	Extreme favour of one preference over the other

Table 15. The definition and explanation scale values

AHP Method - Step 1: Define the alternatives

The AHP Method starts with the definition of alternatives. Alternatives represent the criteria that need to be compared in pairs. The comprehensive list of alternatives is used in the next step.

AHP Method - Step 2: Define the problem and the criteria

The problem breaks into a hierarchy of smaller problems. These problems represent the pairwise comparisons that needs to take place.

AHP Method - Step 3: Establish priority between criteria

The stakeholder establishes the relative importance of a criterion against another one. The comparison between criteria produces a matrix. The elements of the matrix reflect the relative priority between criteria. In case the criterion C_i has higher priority than the criterion C_j for the stakeholder, then the value of P_{ij} is equal to the scale's value, i.e., 3, 5, 7, 9. In case the criterion C_i has lower priority than the criterion C_j for the stakeholder, then the value of P_{ij} is equal to the inverse scale's value, i.e., $1/3$, $1/5$, $1/7$, $1/9$.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
C_1	1	P_{ij}
C_2	...	1	P_{ij}
C_3	1	P_{ij}
C_4	1	P_{ij}
C_5	1	P_{ij}	...
C_6	1	P_{ij}

C_7	P_{ij}	1
	$\Sigma 1$	$\Sigma 2$	$\Sigma 3$	$\Sigma 4$	$\Sigma 5$	$\Sigma 6$	$\Sigma 7$

Table 16. Decision Matrix of AHP Method

AHP Method - Step 4: Determine the weights

The determination of weights requires the normalization of the matrix produced above. Therefore, each array element will be divided with columns' sum (Σ_i), i.e., $P_{ij}' = P_{ij} / \Sigma_j$.

Then, the elements of each row are summed to produce a Priority Vector (P_v) for each row. The weight of each criterion results from the division of Priority Vector (P_v) with the number of criterion (n). Thus, for the criterion C_i , the weight W_i will be;

$$W_i = \frac{1}{n} \sum_{j=1}^n P_{ij}' = \frac{1}{n} * P_v, j = 1, 2, \dots, n$$

Equation 10. Determination of each Preference weight through AHP Method

AHP Method - Step 5: Check consistency of the method

The final step of AHP Method evaluates the consistency of the method. Saaty has established a metric called Consistency Ratio (CR) to evaluate the consistency between the pairwise comparisons as provided by the stakeholder. In other words, the CR index determines whether the AHP Method results are plausible. For example, CR detects if the stakeholder has inserted discrepant preferences, such as extreme importance to minimize the energy cost and extreme importance to maintain the optimal thermal comfort level. The method's results are considered to be sufficiently consistent and acceptable when the $CR \leq 0.1$. In case of higher CR index value, the stakeholder should revise the importance of each preference till the CR index meets the constraint.

The CR index is calculated from the Equation 11 :

$$CR = \frac{CI}{RCI}$$

Equation 11. Determination of Consistency Ratio (CR)

$$CI = \frac{\lambda_{\max} - n}{n - 1}, \text{ where } \lambda_{\max} \text{ is the normalized array's eigenvalue}$$

Equation 12. Determination of each Consistency Index (CI)

Therefore, to calculate the CR index, the values of CI and RCI indexes have to be specified. The CI is calculated from the mathematical formula above (Equation 12), whereas RCI (Randomly Consistency Index) results from a predefined set of values in respect to the number of criteria (n). The values of RI are shown on Table 17.

n	1	2	3	4	5	6	7	8	9
RCI	0	0	0,58	0,9	1,12	1,24	1,32	1,41	1,45

Table 17. The values of RCI for each n-value

AHP Method - Step 6: Conclusions

The ratio between two weights, as determined from the AHP Method, can lead to useful conclusions. It indicates the relative importance of a criterion (Preference) over another one. For instance, if the ratio between the criteria C_5 and C_3 is equal to 2, it means that the criterion C_5 is two times more important than C_3 for the stakeholder.

4.3.2 Multicriteria Decision Making - Vikor method

The **VIKOR Method** is a Multi-criteria Decision Analysis (MCDA) method that applies mostly on problems with conflicting and non-commensurable criteria. Real-life problems usually contain conflicting criteria (e.g., cost - benefit) and non-commensurable criteria (e.g., criteria measured in different units). The VIKOR's Method examines alternative scenarios (A_i) for a family of criteria (C_j) when the weight (W_j) of each criterion is known. The weights can be determined from a method or proposed by the stakeholder's subjective importance evaluation of criteria. The VIKOR's Method produces an output that describes either an ideal solution of one alternative scenario or a compromise solution of two or more alternative scenarios that reaches the ideal one.

Step 1: Define the problem

The definition of the problem requires the specification of the alternative scenarios (A_i), the criteria (C_j), the criteria performances (F_{ij}) and the weights (W_i). As mentioned before, the weight of each criterion can be determined from a method such as EQ, RI, ROC or AHP Method, or proposed by the stakeholder's subjective importance evaluation of criteria.

Step 2: Create the Decision Matrix

The Decision Matrix arranges the alternative scenarios and criteria in rows and columns. alternative scenarios (A_i) are placed in rows and criteria (C_j) are placed in columns. The dimension of the decision matrix is $M \times N$, where M is the number of alternative scenarios and N is the number of criteria. Each element $f(i,j)$ represents the performance of the criterion C_j for the alternative scenario A_i . The performance f_{ij} is an objective and measurable value.

	C_1	C_2	...	C_N
A_1	f_{11}	f_{12}	...	f_{1N}
A_2	f_{21}	f_{22}	...	f_{2N}
...
...
A_M	f_{M1}	f_{M2}	...	f_{MN}

Table 18. The decision matrix of VIKOR's Method

Step 3: Evaluate the Method

The evaluation of the VIKOR Method is based on the L^p -metric which is expressed by the Equation 13.

$$L_i^p = \left(\sum_{j=1}^n \left| \frac{f_j^* - f_{ij}}{f_j^* - f_j^-} \right|^p \right)^{\frac{1}{p}}, 1 \leq p \leq \infty$$

Equation 13. The L^p -metric

For each criterion (C_i), the values of the best-case (f_j^*) and the worst-case scenario (f_j) have to be determined.

If the criterion (C_i) expresses benefit, then;

$$F_j^* = \max \{F_{ij}\} \text{ and } F_j^- = \min \{F_{ij}\}$$

Equation 14. Best-case and worst-case scenario determination for benefit criteria

If the criterion (C_i) expresses cost, then;

$$F_j^* = \min \{F_{ij}\} \text{ and } F_j^- = \max \{F_{ij}\}$$

Equation 15. Best-case and worst-case scenario determination for cost criteria

Step 4: Calculate the metrics S, Q, P.

The S-metric is the weighted L^1 metric and expresses the weighted and normalized distance from the ideal solution (group utility). So, the minimum S-metric value is the optimal one, as it expresses the minimum distance from the ideal solution. For each alternative scenario (A_i), we determine the S-metric value (Equation 16).

$$S_i = \sum_{j=1}^n w_j \left| \frac{f_j^* - f_{ij}}{f_j^* - f_j^-} \right|, \quad i = 1, \dots, m$$

Equation 16. The S-metric

The R-metric is the L^∞ -metric and expresses the highest regret across the criteria (C_j). So, the minimum R-metric value is the optimal one, as it expresses the minimum regret for a specific alternative scenario. For each alternative scenario, we determine the R-metric value (Equation 17).

$$R_i = \max_j w_j \left| \frac{f_j^* - f_{ij}}{f_j^* - f_j^-} \right|, \quad i = 1, \dots, m$$

Equation 17. The R-metric

The Q-metric combines the S-metric and R-metric and can be determined for the Equation 18.

$$Q_i = v \frac{(S_i - S^*)}{(S^- - S^*)} + (1 - v) \frac{(R_i - R^*)}{(R^- - R^*)}$$

Equation 18. The Q-metric

where, S^* expresses the $\min S_i$ and S^- expresses the $\max S_i$

R^* expresses the $\min R_i$ and R^- expresses the $\max R_i$

The constant “ v ” expresses the weight of each metric against the other, i.e., how important is the S-metric against the R-metric. In other words, it weighs the group utility strategy against the minimum regret strategy. Commonly, the two metrics are considered to have the same weight, thus, the v value is 0,5.

Step 5: Determine the optimal solution or the set of compromising solutions.

The problem can have either an optimal solution or a compromising solution. An optimal solution should meet two conditions; the condition of acceptable advantage and the condition of stability. In contrast, a compromising solution meets only one of the two conditions. In case that none of the conditions are met, then, the problem has no solution.

Assume that alternative scenario A_1 has the minimum Q-metric (Q_1) and alternative scenario A_3 follows.

Condition 1; Acceptable Advantage

The alternative scenario A_1 has acceptable advantage in case it meets the condition below (Equation 19).

$$Q_1 - Q_2 \geq DQ \Rightarrow Q_1 - Q_2 \geq \frac{1}{m-1}$$

Equation 19. The required condition to meet acceptable advantage

Condition 2; Stability

The alternative scenario A_1 is stable in case it has both the **minimum S-metric** value and the **minimum R-metric** value.

The alternative scenario A_1 is the optimal solution if it meets the two conditions. Otherwise, the problem has a compromising solution.

- If alternative scenario A_1 meets **only the first condition** of acceptable advantage, then the solution of the problem is the set of alternative scenario A_1 and alternative scenario A_2 which are the two alternative scenarios (A_i) with the **minimum Q-metric values**.
- If alternative scenario A_1 meets **only the second condition** of stability, then the solution of the problem includes all the alternative scenarios (A_i) that meet the condition below (Equation 20).

$$Q_M - Q_1 < DQ \Rightarrow Q_M - Q_1 < \frac{1}{m-1}$$

Equation 20. The required condition to meet acceptable stability

Annexes

A. CDT - Web-application Map

As for now, CDT works with demo data acquired from experiments conducted with a wearable device which developed especially for the W.P. 7.2, data from experiments with a Solar PV Panel, data from a weather REST API. Note that the information shown for the demo user is the minimal information load for development purposes and it not presents data protected by GDPR framework. Additionally, the final authentication mechanism has been described above and it implements an SSO via Keycloak. As the development process adopted is agile, minimal front-end changes may be done if needed or requested by partners. The indicative pages of the CDT described in the previous sections are listed below.

A. CDT - Login Page

Available on: <https://www.consumerdigitaltwin.com/login.php>

B. CDT - Dashboard

Available on: <https://www.consumerdigitaltwin.com/index.php>

C. CDT - Energy Production

Available on: https://www.consumerdigitaltwin.com/energy_production.php

D. CDT - Preferences

Available on: <https://www.consumerdigitaltwin.com/preferences.php>

E. CDT - Clothing Insulation

Available on: https://www.consumerdigitaltwin.com/clothing_insulation.php