D4.1 SmartLivingEPC CIEM & Building Dynamic Behaviour Monitoring Platform

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Project Acronym:	SmartLivingEPC
Project Full Title:	Advanced Energy Performance Assessment towards Smart Living in Building and District Level
Grant Agreement:	101069639
Project Duration:	36 months (01/07/2022 – 30/06/2025)

D4.1 SmartLivingEPC CIEM & Building Dynamic Behaviour Monitoring Platform v1

Work Package:	WP4-Building digitalization towards Smart Living EPCs
Task:	T4.1 Common Information Exchange Model for Smart Living EPCs T4.2 Building dynamic behaviour monitoring (both energy and non-energy)
Document Status:	Final
File Name:	SmartLivingEPC _ D4.1_SmartLivingEPC CIEM & Building Dynamic Behaviour Monitoring Platform
Due Date:	30.11.2023
Submission Date:	22.12.2023
Lead Beneficiary:	CERTH

Dissemination Level	
Public	☒
Confidential, only for members of the Consortium (including the Commission Services)	☐

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Authors List

Leading Author				
First Name		Last Name	Beneficiary	Contact e-mail
Samy		Iousef	QUE	s.iousef@que-tech.com
Babis		Bazakas	CERTH	chbazakas@iti.gr
Aggeliki		Veliskaki	CERTH	aveliskaki@iti.gr
Co-Author(s)				
#	First Name	Last Name	Beneficiary	Contact e-mail
1	Beatriz	Fraga De Cal	IES	beatriz.fraga@iesve.com

Reviewers List

Reviewers			
First Name	Last Name	Beneficiary	Contact e-mail
Stergios	Kokorotsikos	EUNICE	skokorotsikos@eunice-group.com
Paul	Waide	WSEE	paul@waide-europe.eu
Corin	Waide	WSEE	corin@waide-europe.eu

Version History

v	Author	Date	Brief Description
0.1	Samy Iousef (QUE)	15 October, 2023	Initial Draft (ToC)
0.2	Babis Bazakas (CERTH)	16 October, 2023	Modifications on ToC
0.3	Samy Iousef, Babis Bazakas (QUE, CERTH)	29 November, 2023	First Draft of the content
0.4	Aggeliki Veslikaki (CERTH)	30 November, 2023	Comments and content addition
0.5	Samy Iousef, Babis Bazakas, Aggeliki Veliskaki (QUE, CERTH)	18 December, 2023	Final Draft for internal review
0.6	Dimosthenis Ioannidis (CERTH)	19 December, 2023	Quality check and minor corrections

0.7	**Samy Iousef, Babis Bazakas, Aggeliki Veliskaki (QUE, CERTH)**	**21 December, 2023**	Updated based on comments from reviewers
0.8	**Aggeliki Veliskaki (CERTH)**	**21 December, 2023**	Available for submission to the EC

Legal Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA) or the European Commission (EC). Neither the European Union nor the granting authority can be held responsible for them.

Copyright

Executive Summary

This deliverable is a joint presentation of the work carried out under T4.1 Common Information Exchange Model (CIEM) for Smart Living EPCs and T4.2 Building dynamic behaviour monitoring (both energy and non-energy). The first aim of this deliverable is to provide a detailed description of the CIEM undergoing development in the SmartLivingEPC project, with a focus on its planned architecture, sub-component functionality, and data schemas. The deliverable also provides a detailed description of the development and evaluation processes of the SmartLivingEPC's Building dynamic behaviour system.

The SmartLivingEPC framework handles extensive static and real-time data gathered from diverse sources within the building. All this information needs to be transformed into understandable formats for other SmartLivingEPC components. Within SmartLivingEPC, the CIEM holds comprehensive knowledge related to the building, encompassing 3D maps, insulation, assets, BIM, audits, SRIs, energy sources, context details, and sensors, etc. Additionally, it involves developing an IoT sensor middleware to facilitate real-time data exchange among SmartLivingEPC components. Furthermore, the project introduces the Building Dynamic Behaviour Monitoring System, responsible for modelling and implementing the building's dynamic behaviour. It considers dynamic building data such as energy usage, indoor conditions, and occupancy, correlating and organizing this information to derive the model for the building's behaviour. This approach focuses on the "occupant" parameter, recognizing its significant influence on the building's overall behaviour.

This document provides a framework for SmartLivingEPC interoperability, emphasizing the integration of Building Information Modelling (BIM) and Internet of Things (IoT) data. It explores methodologies for harmonizing these data streams to establish a cohesive ecosystem. Central to this framework is the SmartLivingEPC CIEM system, acting as a data collection, management and sharing system. It encompasses a robust data model, management strategies, and a multi-layered architecture supported by a defined technology stack. Additionally, it outlines the connections with pilot sites.

The document also addresses literature review and data acquisition specifics for Building Dynamic Behaviour Monitoring System. It enables a deeper understanding of occupancy patterns within buildings and details load categories and temporal granularity for effective data collection. Moreover, it explores the methodology and algorithms employed in the Building Dynamic Behaviour Monitoring System. This includes analytics, modelling techniques, and behaviour generators designed to interpret and simulate dynamic building data.

The deliverable details the actual implementation of the project component up to M18, taking into account the current development state under a public dissemination level.

The second and final version of this deliverable, due by M22 of the project is expected to demonstrate the entire design and implementation of the SmartLivingEPC CIEM, in conjunction with the completed development of Building Dynamic Behaviour Monitoring System.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Table of Contents

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

List of Figures

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

List of Tables

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

List of Acronyms and Abbreviations

Term	Description
AAI	Authorization Infrastructure
AEC	Architecture, engineering, and construction
AI	Artificial intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
BIM	Building Information Modelling
BIMRL	BIM Rule Language
BMS	Building Management Systems
BOT	Building Topology Ontology
BPD	BIM Perspective Definition
CARTree	Classification And Regression Tree (CART)
CDBLSTM	Convolutional Deep Bi-directional Long Short-Term Memory
CIEM	Common Information Exchange Model
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DCV	Demand-Controlled Ventilation
DT	Decision Trees
DTC	Decision Tree Classifier
ELM	Elaboration likelihood model
EPC	Energy Performance Certificate
ExKMC	Explainable K-Means Clustering
FM	Facility Manager
GB	Gradient Boosting
HMM	Hidden Markov Model
HVAC	Heating, Ventilation and Air-Conditioning
IDS	Information Delivery Specification
IFC	Industry Foundation Classes
IoT	Internet of Things

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

KNN	**K-Nearest Neighbors**
LBD	**Linked Building Data**
LDA	**Linear Discriminant Analysis**
LR	**Logistic Regression**
LSTM	**Long Short-Term Memory**
ML	**Machine Learning**
MVD	**Model View Definition**
NNs	**Neural Networks**
NRMSE	**Normalized Root Mean Square Error**
OE	**Occupancy Estimation**
OWL	**Web Ontology Language**
RDF	**Resource Description Framework**
RF	**Random Forest**
RVT	**Revit**
SAREF	**Smart Applications REFerence Ontology**
SOSA	**Sensor, Observation, Sample and Actuator**
SRI	**Smart Readiness Indicator**
SSN	**Semantic Sensor Network**
SSO	**Single Sign-On**
SV	**Support Vector**
SVM	**Support Vector Machines**
W3C	**World Wide Web Consortium**
XGBoost	**eXtreme Gradient Boosting**

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

1 Introduction

1.1 Scope and objectives of the Deliverable

WP4 aims to create the SmartLivingEPC's CIEM and an IoT platform to assist in visualizing SmartLivingEPC data and aiding decision-making about building performance. It will also develop a digital twin tool and design a digital logbook for recording data and actions. This WP will also provide the EPC calculation module and its corresponding APIs to ensure the independency of each module.

The scope of deliverable D4.1: *SmartLivingEPC CIEM & Building Dynamic Behaviour Monitoring Platform*, is to present the design and implementation details of the Common Information Exchange Model (CIEM) and the Building Dynamic Behaviour Monitoring System that are developed in terms of WP4. The CIEM is a data model that defines the structure and format of the information that is exchanged among the different components of the SmartLivingEPC project. The Building Dynamic Behaviour Monitoring tool is a system that collects and processes dynamic building information from various sources, such as sensors and smart meters to monitor the dynamic behaviour of buildings. This deliverable reports the results of both T4.1 and T4.2, which are focused on the harmonization, alignment and establishment of the appropriate tools and procedures for data collection, interoperability, and monitoring.

More specifically, T4.1: *The SmartLivingEPC CIEM* is responsible for managing and integrating various types of data that are relevant for the building's performance and sustainability. The data sources include both static and dynamic information, such as insulation materials, assets and equipment, Building Information Modelling (BIM), audits, Smart Readiness Indicator (SRI), energy and non-energy sources, context information (e.g. energy consumption, generation and storage, non-energy sources consumption etc.), sensors, etc. The SmartLivingEPC CIEM will provide a common representation of the data that can be understood and used by other SmartLivingEPC components. Additionally, this task will develop an IoT sensorial middleware for real-time data exchange among the various SmartLivingEPC components.

In the framework of T4.2: *The Building Dynamic Behaviour Monitoring System* the aim is to create and implement a model that captures the dynamic behaviour of the building. The model will consider various dynamic building data (such as energy use, indoor environment, occupancy, etc.) and will analyse how they relate, integrate, and categorise them to estimate the existence of occupants in the building and identify the profile of occupant's behaviour. The model will be user orientated, since the "occupant" variable/ activity is the main factor that influences the overall behaviour of the building.

1.2 Relation to other Tasks and Deliverables

This deliverable has a strong relation with other tasks and deliverables of the SmartLivingEPC project, which should be considered along with this report for a proper understanding of its contents.

In Task T1.3 *Pilot Surveys and Use Case Scenarios definition,* the scenarios' requirements were determined by conducting a survey on the pilot sites. The survey report evaluated the current state of the pilots in terms of energy performance, building structure, well-being (indoor comfort rating), user experience, facilities, and infrastructure. The information collected enabled the assessment at the complex scale, considering the neighbourhood infrastructure, smart grids and other energy community practices. The pilot buildings constitute various use-case scenarios based on the descriptions of the competences and impact for each of the 9 demonstration surveys. Results and findings from T1.3 fed into T1.4 *Technical requirements & architecture design*, which provides a comprehensive technical description of the SmartLivingEPC solution and the specifications for each of its key components, modules, and their functionalities. In this task the overview of the SmartLivingEPC solution was defined as the Conceptual Architecture Design, its components and sub-components, their interfaces and the connections with external systems (e.g., legacy systems for data provision, BIM, IoT sensor networks). Furthermore, the outcomes of T1.3 and T1.4, included in D1.2, contributed to the definition of the Modules Functional and Technical Specifications. This part has two objectives, firstly to provide a high-level sketch of dependencies among the various interrelated modules of the SmartLivingEPC solution (software and technology solutions) and secondly to describe in detail the functional decomposition of each sub-system. The constraints of the architecture elements/layers in terms of software and/or hardware resources, compatibility with standards, interoperability, and all types of data flowsheet. These aspects enabled the development of a clear integration roadmap for the project's tools.

The relation of D1.2 SmartLivingEPC pilot analysis, use case scenarios and Framework Architecture including the results of T1.3 & T1.4 has a strong relation with D4.1 SmartLivingEPC CIEM & Building Dynamic Behaviour Monitoring Platform as they are part of the SmartLivingEPC solution for energy efficiency and comfort optimization in buildings. The pilot analysis and use case scenarios define the scope and objectives of each pilot site, while the Framework Architecture provides the overall structure and functionality of the SmartLivingEPC solution. The SmartLivingEPC CIEM & Building Dynamic Behaviour Monitoring System are core components of the SmartLivingEPC framework since they are used as intermediate components for the successful utilization of Smart Living EPC services.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

1.3 Structure of the Deliverable

As presented in Section 1, this deliverable contains the outcome of two different tasks, T4.1 and T4.2, which have been conducted in parallel to jointly respond to the project's initial requirements.

Section 1 introduces the scope and objectives of this deliverable, its relation to other tasks and deliverables in the project, and its structure. Section 2 presents an extended review of the literature on the integration of BIM and IoT data, focusing on the challenges and benefits of the existing frameworks and standards that support them. Section 3 provides a comprehensive assessment of the literature on occupancy estimation and occupancy profiling approaches. Section 4 presents the CIEM data model and CIEM data management system, explaining their design principles, components, and functionalities. In addition, information regarding the interconnection of CIEM with pilot sites is also provided. Section 5 presents the data that are utilized in Building Dynamic Behaviour Monitoring systems for model creation. Section 6 outlines the methodology employed for monitoring the dynamic behaviour of buildings. This section details also the techniques and tools used to analyze collected data and the design process to generate building dynamic behaviours. Section 7 introduces the algorithms utilized within the dynamic behaviour monitoring system, describing the methods used to clean, transform, or prepare raw data before analysis. This section specifically addresses the training and evaluation process for models used in estimating occupancy and in occupancy profiling within the building. Section 8 outlines the integration of the project's outcomes with the SmartLivingEPC framework, discussing the role of CIEM in this process and plans for enhancing the Building Dynamic Behaviour Monitoring Platform and CIEM. Furthermore, Section 8 discusses the next steps for the implementation and validation of the CIEM and the Building Dynamic Behaviour Monitoring Platform.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

2 SmartLivingEPC interoperability framework

2.1 Integration of BIM and IoT data

The architecture, engineering, and construction (AEC) industry has seen a rapid digital transformation in recent years. In particular, with the integration of cutting-edge technologies like BIM [1] and Internet of Things (IoT), the AEC sector has become more efficient, and collaborative in a more sustainable environment. These digital tools have revolutionized project design and management, enabling real-time evaluation, advanced modelling and simulations and, eventually, data-driven decision-making. As a result, new construction and renovation projects are now completed faster, with fewer errors, and at a reduced cost, ultimately leading to more sustainable, energy efficient and innovative built environments.

BIM is an advanced and collaborative method for designing and managing construction and infrastructure projects, involving open standards and workflows that enhance the interoperability among stakeholders and eventually the project efficiency [2]. It involves creating and maintaining digital representations of a project, spanning from initial planning to construction and ongoing maintenance, supporting the design of new building [3] or the renovation of existing ones [4]. A plethora of advantages accompany the use of BIM, a selection of which are now outlined. By minimizing errors and discrepancies, BIM results in cost savings and superior project outcomes. Additionally, it empowers better decision-making and facilitates more informed choices about design modifications and material selections. BIM also supports sustainability efforts by enabling the analysis of energy usage and environmental impact, making it an essential tool in the contemporary construction industry. In essence, BIM transforms the way projects are conceived, executed, and managed, ushering in a new era of efficiency, precision, and sustainability in construction and infrastructure development.

Despite its advantages, BIM implementations/solutions often make use of proprietary file formats that are tailored to the needs of certain vendors (e.g., Revit[1] developed by Autodesk, uses its proprietary Revit (RVT) file format for storing BIM data). Using proprietary files hinders the ability to share data across different software platforms and potentially creates interoperability issues when various stakeholders are involved in the project's process. To help to address this, openBIM arises as a concept that outlines the use of open standards and workflows. As opposed to BIM, openBIM promotes the use of non-proprietary, open-source software and file formats that are not tied to a particular vendor. openBIM's focus lies on interoperability, enabling different tools to exchange information seamlessly and enhancing collaboration among the various stakeholders [3]

A widely used openBIM standard is the Industry Foundation Classes (IFC) schema developed by buildingSMART International, a consortium of construction and infrastructure professionals, software vendors, and academic institutions[2]. In line with the BIM core aspects, the IFC schema stands as an open and neutral data format standard within the AEC industry, serving as a digital representation framework, facilitating interoperability and data exchange among various stakeholders.

The IFC schema enables the seamless data sharing of building-related information models across different software vendors thereby fostering collaboration, reducing errors, and improving efficiency throughout the lifecycle of a building project. The IFC specifications are formatted based on the STEP ISO standard, either in eXtensible Markup Language (XML)[3] or in EXPRESS[4]. There are several available IFC EXPRESS schemas, e.g., IFC2X3, IFC2X3 TC1, IFC4 and IFC4 ADD1.

[1] https://www.autodesk.com/products/revit/architecture

[2] https://technical.buildingsmart.org/standards/

[3] https://www.iso.org/standard/40646.html

[4] https://www.iso.org/standard/38047.html

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

With IFC being an extensive information schema, covering domains related to architecture, structure, mechanical, electrical, plumbing, and civil engineering etc., it often requires that simplifications are made to (i) align with the requirements imposed by specific use cases and (ii) maintain the exchanged information within the desired frames [5]. To do so, researchers and practitioners have resorted to the development of Model View Definition (MVD) (e.g., [6]). MVD constitutes a subset of the IFC schema containing only specific elements based on a defined exchange requirement[5]. By making use of MVDs, it is possible to validate the completeness of a BIM model against predefined concepts via the mvdXML specification. An example of such a validation exercise is depicted in Figure 1. In this case, completeness checking is made to ensure that each building element of the IFC model has a property used to create a semantic link between the element and its thermal properties.

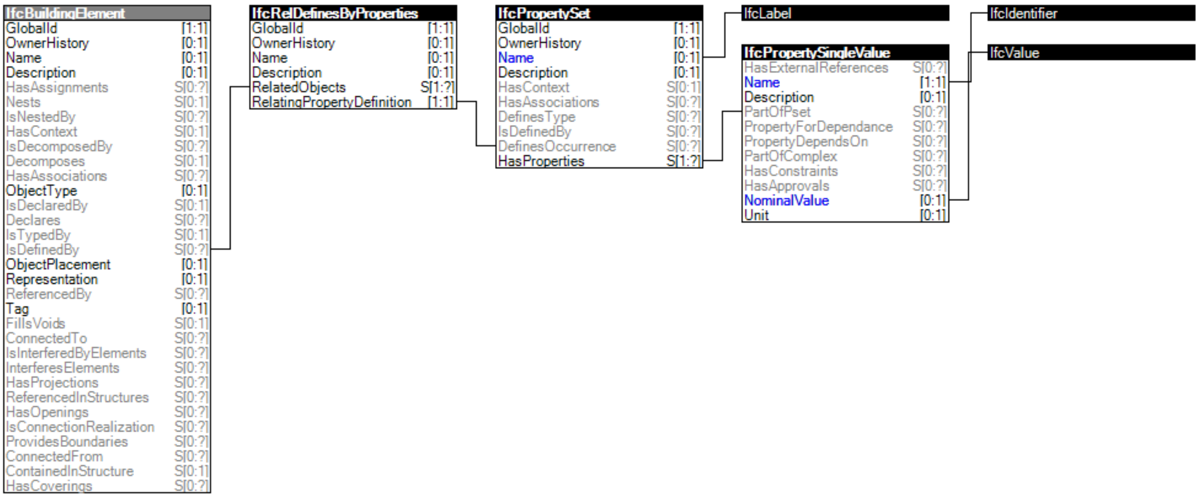

Figure 1: MVD concept template for IFC validation

A particular concern that has emerged in recent years is the vast number of MVDs created for various software vendors, leading to interoperability issues. This, eventually, necessitates additional efforts for software implementation. To overcome the aforementioned challenge, buildingSMART's efforts concentrate on the development of a standard for specifying and checking simple information requirements from IFC models, namely the Information Delivery Specification (IDS) standard.

The transition from MVD to IDS implementations is primarily driven by the need for enhanced interoperability within software tools[5]. The multiple MVDs previously in use lacked interoperability among themselves, requiring extensive efforts for implementation across software platforms. Therefore, this transition aims to streamline the process by establishing a standardized approach through IDS, enabling more seamless compatibility between various levels of IFC implementation and fostering greater interoperability across domains. The IDS standard, working alongside IFC and in line with the ISO 29481-1[6], facilitates the definition of computer-interpretable exchange requirements tailored to specific use cases, further promoting smoother data exchange and collaboration within the industry. IDS tools (file creation and checking) are provided by many software vendors allowing any user to develop customised IDS Specifications from scratch or modify an existing one from a public IDS template. Eventually, any IFC file can be checked by an IDS file, irrespective of the software that was used to produce it[7].

[5] https://technical.buildingsmart.org/standards/ifc/mvd/

[6] ISO TC 59/SC 13, ISO 29481-1: 2016, Building information models — Information delivery manual — Part 1: Methodology and format, ISO, Geneva, Switzerland, 2016.

[7] https://github.com/buildingSMART/IDS/tree/master/Documentation

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Complementing the use of BIM in the AEC industry, the integration of IoT represents a pivotal shift in the way buildings are designed, constructed, and, eventually, operate. At its core, IoT in buildings involves the deployment of a vast network of interconnected sensors and devices (e.g., [7]) capable of monitoring and controlling various aspects of the building's environment. These sensors can collect and process in real-time a wide array of data, including, among others, temperature, humidity, light levels, air quality, occupancy etc., allowing thus for informed decisions while establishing a high level of automation. The aforementioned points are particularly important in the case of Facility Managers (FMs) who need to continuously monitor the performance of buildings and proactively proceed with maintenance tasks.

The fusion of static BIM data and dynamic information received from the IoT sensors emerges as a critical need in the building sector to provide the involved stakeholders with comprehensive and real-time insights throughout a building's lifecycle [8]. By combining BIM, which provides a detailed digital representation of a building's design and structure, with IoT devices that gather real-time data from sensors embedded within the building's systems, stakeholders gain a holistic view of building performance. This fusion enables enhanced monitoring, predictive maintenance, and optimization of energy usage, occupant comfort, and overall operational efficiency. Integrating BIM and IoT data empowers stakeholders to make data-driven decisions, streamline maintenance, detect issues proactively, and create smarter, more sustainable buildings that adapt to evolving needs while maximizing efficiency and performance [9] [10]. Within SmartLivingEPC, the abovementioned fusion certainly raises the need for an efficient data management framework that, among others, serves the SmartLivingEPC architecture, currently documented in the *D1.2 - SmartLivingEPC pilot analysis, Use case scenarios and Framework Architecture v1*. In the following section, scrutinization of the available literature is performed in order to explore the available BIM-IoT data integration methods and identify the optimal one that shall constitute the SmartLivingEPC solution.

2.2 BIM and IoT data integration frameworks

Comprehensive reviews on the available BIM-IoT data fusion frameworks are documented in the available literature [8] [11] [12], providing comparative analysis of the methods currently employed towards a digital transformation, enhanced interoperability and efficiency of the AEC industry services. Among the reviewed methods, and in line with the relevant work by Tang et al. (2019) [11], the following methods are noted as potential methods to be used in SmartLivingEPC:

- combining Application Programming Interfaces (APIs) of the BIM authoring tools along with relational database;
- transforming BIM data into relational database based on a new data schema and;
- employing semantic web technologies.

Note that, also other methods are noted in the reviewed literature, such as the creation of a new query language, hybrid methods etc. However, within the (time-) frames of the SmartLivingEPC project, these are considered as less relevant and therefore not accounted for in the present analysis.

The former method, i.e., the integration of BIM tools with sensor data leveraging the tools' APIs and relational databases, involves storing sensor-collected time-series data in a relational database while exporting BIM models into compatible database formats through Application Programming Interfaces (APIs). Establishing a clear database schema facilitates the connection between virtual objects in the BIM models and physical sensors. Examples from various studies, such as Marzouk et al. (2014) [13] and Zhang and Bai (2015) [14], demonstrate the utilization of this method to link Revit models with sensor data stored in databases like Microsoft Access or MySQL. While offering advantages, such as ease of integration and automatic data updates, this approach is best suited for simpler BIM models with fewer sensors [11]. In addition, it presents limitations in parameter exports and requires manual model file updates upon changes. Despite drawbacks, its accessibility and compatibility foster wider adoption and ease in linking model and sensor data, especially when integrating existing Building Management Systems (BMS) with BIM.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

The second approach addressing the integration of sensor data with BIM involves transforming BIM data into a queryable database structure, enabling information extraction from various user perspectives. This method involves the conversion of BIM data into a relational database structure, which facilitates the linkage between time-series sensor data and BIM elements. Examples of this approach are included in the study by Solihin et al. (2017) [15]. In their study, the authors presented a methodology to transform IFC-related data to a new schema, namely the BIMRL schema, allowing efficient SQL queries on BIM data without storing complete IFC-STEP data. Kang and Choi's work [16] aimed to bind BIM and facility management using a BPD metadata structure, enabling data extraction based on user queries. While this approach offers flexibility in expanding user perspectives and leveraging existing SQL, creating a new data schema demands substantial effort in data mapping and expertise in various domains like language design, IFC, databases, and programming [11]. Despite the manual [14] [13] construction of virtual objects for sensors, this method suits complex projects with intricate spatial contexts and numerous sensors, as it tailors the schema based on user perspectives rather than exporting entire complex IFC data.

In recent times, there has been an increasing interest in research focusing on the possibilities of connecting and merging cross-domain building data i.e., BIM and IoT, through semantic web technologies [17]. A well-established representation of the Industry Foundation Classes is the ifcOWL that constitutes a Web Ontology Language (OWL)[8]. As noted by Pauwels and Terkaj [18], ifcOWL enables the seamless use of the well-established IFC standards for construction data representation, leveraging the advantages of semantic web technologies for data distribution, flexible data model extension, efficient querying, and logical reasoning. Additionally, it facilitates the utilization of general-purpose software implementations for tasks such as data storage, ensuring consistency, and performing knowledge inference. On the other hand, the Semantic Sensor Network (SSN) is an ontology within the domain of the semantic web designed to describe and represent sensor-related data (sensor features and properties, procedures, samples, actuators etc.) in a semantic, machine-understandable format. It [15] is based on a modularized architecture and the SOSA (Sensor, Observation, Sample and Actuator) ontology [19] aiming at providing a standardized way to describe, publish, discover, and integrate heterogeneous sensor data.

Semantic web technologies offer a promising solution for handling intricate, interconnected data across diverse domains, particularly addressing challenges in cross-domain collaborations with multiple data schemas. By standardizing how information is conceptualized and linked, these technologies enable seamless search, composition, and translation of data across disciplines. Successful implementations of such technologies have been documented within the framework of H2020 projects such as BIMERR[9] and Vicinity[10]. Furthermore, in the past, researchers (e.g., [20]) have constructed semantic models that integrate domain-specific rules for reasoning. In this case, Resource Description Framework (RDF) format is utilized for environmental sensor data and building information (e.g., [21]), employing SPARQL rules to detect anomalies in compliance with building regulations. RDF offers a standardized approach for outlining web resources through a graph-oriented data structure. Every detail is articulated as a triplet, comprising a subject, predicate, and object, forming the core structure of information representation.

Despite their advantages, semantic techniques encounter certain challenges such as: (i) data format translation hurdles between BIM and IoT; (ii) the need for multiple ontologies to address complex problems; (ii) inference slowdowns with large ontologies etc. Additionally, there are limitations in fully integrating a semantic dataspace and difficulties in handling dynamic time-series data from IoT devices. Nevertheless, these technologies notably resolve interoperability issues at a semantic level for BIM and IoT, showcasing their potential, albeit the need for further exploration in leveraging dynamic time-series data in the IoT infrastructure.

In line with the reviews by Tang et al. (2019) and Huang et al. (2023), the use of semantic web technologies constitutes the most promising method for the delivery of a relevant overview of well-defined building-related data, that represents both the BIM and IoT domain information with semantic interoperability. In this case,

[8] https://technical.buildingsmart.org/standards/ifc/ifc-formats/ifcowl/

[9] https://bimerr.eu/

[10] https://vicinity2020.eu/

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

leveraging the advantages of semantic web technologies, provide a common data environment that concentrates, manages and shares data from diverse domains, fostering data exchange among different software systems and stakeholders. To this end, and within the frames of the SmartLivingEPC project, the use of semantic web technologies is further explored here, noting hereafter the most relevant initiatives, namely the Brick schema, BOT ontology and Linked Building Data (LBD) as well as the Smart Applications REFerence Ontology (SAREF) [11]. SmartLivingEPC will leverage the advantages of these initiatives to develop the SmartLivingEPC solution for a fused BIM and IoT data management.

The Brick schema is a pivotal open-source initiative aimed at standardizing semantic descriptions of physical, logical, and virtual assets within buildings, along with their interconnected relationships [19]. It comprises an extensible dictionary of building-related terms, a network of relationships for linking and composing these concepts, and a flexible data model that seamlessly integrates with existing tools and databases. Leveraging the resource description framework, Brick adopts a subject-predicate-object model, enabling robust relations between building components. Through powerful Semantic Web technology, Brick ensures a consistent description of diverse building features and subsystems. Its adoption as the canonical building description offers numerous advantages: reducing deployment costs for analytics and energy efficiency measures, providing an integrated representation of various subsystems like HVAC, lighting, fire, and security, simplifying development of smart applications, and diminishing reliance on non-standard labels prevalent in building management systems. Notably, Brick is freely available under the BSD 3-Clause license[12], encouraging widespread adoption and collaborative development via its GitHub repository[13]. In Figure 2 an example of a basic Brick model is illustrated.

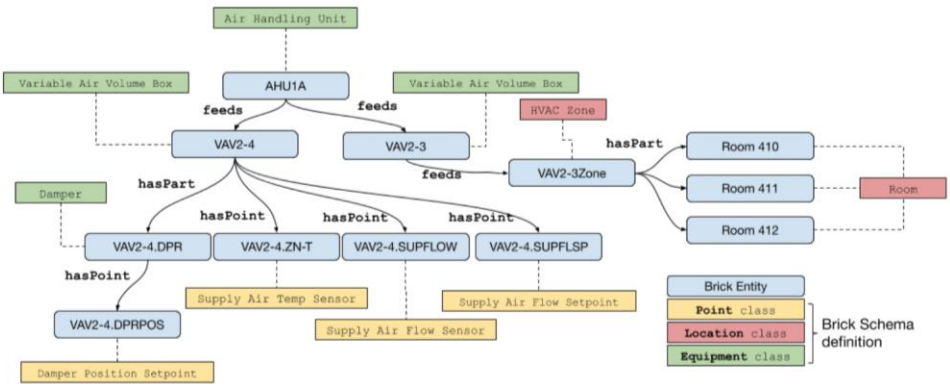

Figure 2: Example of Brick dataset[14]

The Linked Building Data community group is an initiative formed in the World Wide Web Consortium (W3C)[15] that aspires to advance the use of linked data principles and web technologies within the AEC industry. It brings together all the relevant stakeholders, including researchers, developers, industry professionals, and organizations, with the aim of exploring and promoting the application of semantic web technologies in the domain of construction and building information, by encouraging the adoption of standardized data formats and ontologies, such as the Building Topology Ontology (BOT) for representing building information in a machine-readable and interconnected manner. BOT, being an ontology that represents the topological information of

[11] https://saref.etsi.org/

[12] https://github.com/BrickSchema/Brick/blob/master/LICENSE

[13] https://github.com/BrickSchema/brick/

[14] https://brickschema.org/

[15] https://www.w3.org/community/lbd/

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

buildings, can be structured and defined using RDF to describe the various elements, relationships, and properties within a building's topology [19].

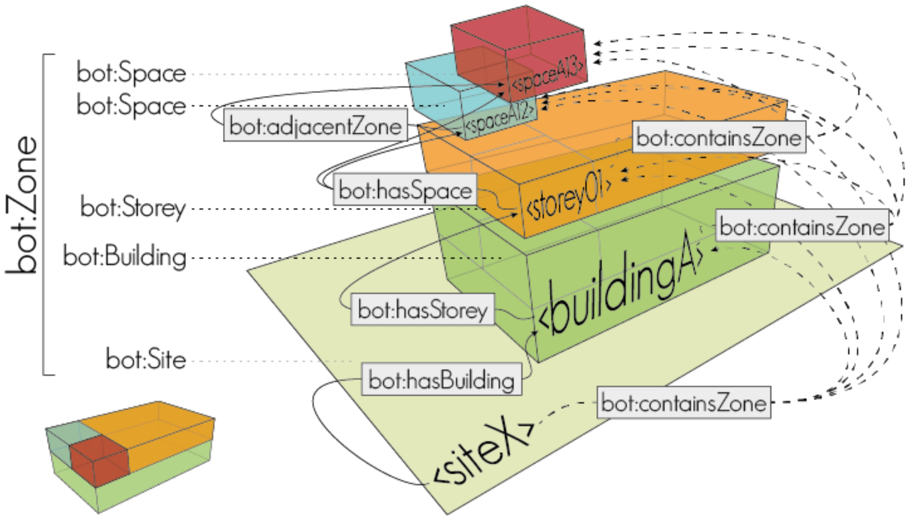

Figure 3: BOT related classes and relationships involved in Zones[16]

The SAREF stands as a modular and versioned ontology framework, fostering semantic interoperability among diverse IoT solutions across industries [22]. Originating from a team at TNO (Toegepast Natuurwetenschappelijk Onderzoek) in the Netherlands, SAREF has evolved into a shared consensus model adept at aligning existing standards, protocols, and data models in the realm of smart appliances. It offers modular building blocks, enabling the segregation and fusion of ontology components based on specific application requirements. SAREF is standardized and maintained under the European Telecommunications Standards Institute Specialist Task Force (ETSI STF) and other University partners/contributors who have enriched and refined SAREF along with insights received from technical and domain expert groups. This collaborative effort synthesizes knowledge from diverse research and development projects. In the AEC industry, SAREF's adaptable structure and consensus-driven model facilitate seamless integration and interoperability among various smart appliances, optimizing processes and workflows across the sector. SAREF's core ontology along with its domain extensions are currently freely available on a local GitLab account[17].

[16] https://w3c-lbd-cg.github.io/lbd/bot/

[17] https://labs.etsi.org/rep/saref

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Serving as an extension of the SAREF ontology, the SAREF4BLDG[18] module aims at broadening the range of concepts, properties and relationships that are relevant to buildings ensuring a more accurate and detailed representation of building-related data within the semantic web framework. Furthermore, SAREF4BLDG's development takes into account the Industry Foundation Classes standard for building information, and in particular the IFC classes and properties related to smart appliances [19]. Figure 4 depicts an overview of the SAREF4BLDG extension classes (only the top levels of the hierarchy) along with the relevant properties.

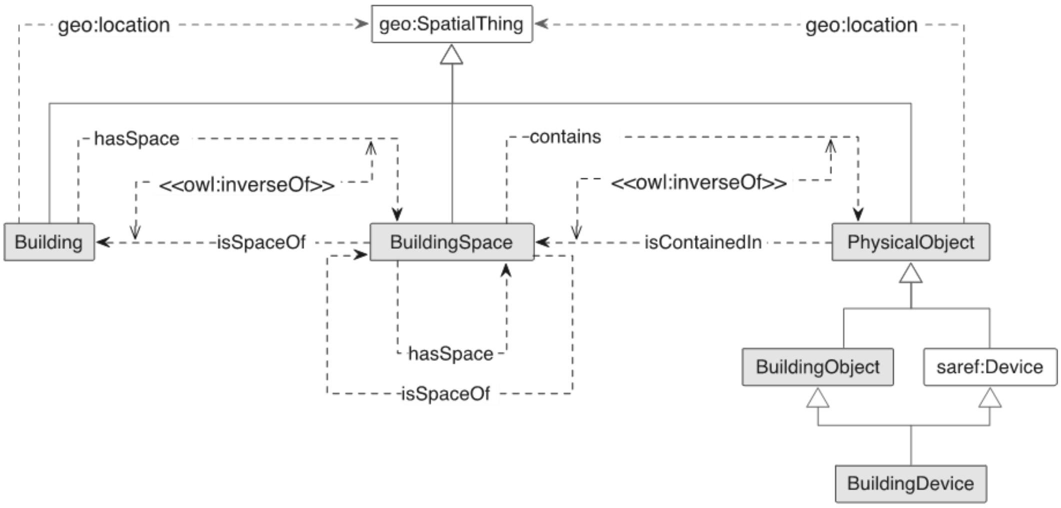

Figure 4: General overview of the top levels of SAREF4BLDG[19]

[18] https://saref.etsi.org/saref4bldg/v1.1.2/#[1]

[19] https://saref.etsi.org/saref4bldg/v1.1.2/

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

3 Data utilization in AI and ML services

Based on the above, an extensive repository that contains diverse building data sourced from BIM, simulation outcomes, or IoT data streams can be developed. Advancements in AI computational power facilitate the enhanced analysis of this dataset, enabling precise predictions and decision-making procedures [23] [24]. Utilizing this potential for in-depth analysis of building performance can foster sustainable building administration, heighten occupants' consciousness about energy usage, and reduce environmental footprints [25].

AI models have primarily found application in the building sector concerning the building's energy usage. Initially, AI algorithms are employed to forecast diverse energy demands within buildings, like heating or cooling needs, aiding in more effective systems management. By integrating Machine Learning (ML) tools with IoT controllers, intelligent temperature regulation based on occupants' preferred temperatures can be achieved, consequently reducing overall building energy usage [26]. Another category of AI algorithms focused on buildings' energy consumption assesses construction designs and offers predictive insights, suggesting retrofitting actions for achieving the most efficient design to minimize energy consumption [27]. The incorporation of AI is anticipated to become indispensable during both the design and operational phases of energy-efficient construction in the foreseeable future [28].

Moreover, AI cooperation with IoT data supports the identification of occupancy behaviour in a building which can have significant impact on the energy performance of a building and the operation of installed systems such as heating, ventilation and air-conditioning (HVAC) systems [29]. Furthermore, the profiling of the occupants behaviour in different sectors of buildings could be utilized in order to improve the operation of HVAC systems, increasing the energy efficiency of the building [30] [31].
As already mentioned, T4.2 aims to develop a Building dynamic behaviour monitoring system focusing on occupancy parameter. The goal is to identify occupancy presence and estimate occupancy profiles for use in various applications within the SmartLiving EPC framework. The data gathered and correlated in terms of CIEM will be utilized along with AI models in order to cover the requirements of T4.2. Building monitoring system – Occupancy related.

3.1 Occupancy estimation approach

Occupancy estimation within buildings has emerged as a pillar in the pursuit of energy efficiency and optimized functionality. A variety of sensors, including environmental sensors, has been essential in this realm, offering non-intrusive and cost-effective solutions. Traditionally, the process involved manual feature extraction, demanding strong domain expertise and often missing implicit yet valuable features. Recent advances in sensor fusion have resulted in a variety of systems for occupancy estimation and detection. From passive infrared sensors to cameras and CO2 sensors, each offers unique capabilities and constraints, leading to a categorization of solutions based on environmental sensors [32].The integration of multiple sensor types has been contributing to more comprehensive occupancy estimations and detections.
Studies have highlighted the impact of occupancy on energy consumption within buildings. Efforts such as occupancy-based HVAC scheduling have demonstrated energy savings upwards of 9%, with more sophisticated predictive control algorithms highlighting reductions of around 40% in energy usage [33]. Moreover, the implications extend beyond energy conservation, encompassing lighting systems, security management, and emergency protocols. However, while commercial buildings have seen significant attention in occupancy modelling, the domain's focus on residential buildings has been comparatively limited. This gap necessitates unique models catering to the requirements of residential settings, considering privacy concerns and the variability in occupancy states [34]. Developing predictive models specifically tailored for residential settings and different temporal and spatial resolutions can significantly support energy-efficient controls in diverse building environments.

Extensive research has been carried out to determine occupancy in a building. These methodologies are broadly classified into four principal types: classic statistical strategies, unsupervised machine learning techniques, supervised machine learning techniques, and mixed or hybrid machine learning approaches.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

The studies that align with the traditional statistics model are those based on Markov chain models [35], [36]. To forecast both the binary-level occupancy and the actual count of occupants in U.S. office spaces, Li and Dong [35] developed a Markov model that incorporates change-point analysis, using a moving window for training and an adapted random sampling technique. This model was benchmarked against two simulation approaches (Page's Markov and Reinhart's Lightswitch) and two machine learning algorithms (Artificial Neural Network (ANN) and Support Vector (SV) regression). The effectiveness of these methods was evaluated based on their ability to predict occupant presence and numbers with a lead time of 15 minutes, 30 minutes, and 24 hours. The findings demonstrated that Li and Dong's Markov model surpassed the performance of the other four methodologies considered in predicting occupancy. Additionally, [36] introduces an innovative occupancy estimation method for buildings, employing infrared array sensors and advanced machine learning techniques, including Non-negative Matrix Factorization and Inhomogeneous Hidden Markov Models, to accurately determine indoor occupancy levels, significantly enhancing the efficiency of building Heating, Ventilation, Air Conditioning, and Lighting systems.

The contemporary literature on occupancy prediction increasingly favours the application of machine learning techniques, owing to their adaptability and precision in forecasting occupancy, whether in binary terms or by quantifying the number of occupants.

ML algorithms are commonly categorized into supervised and unsupervised learning methods. In unsupervised learning, models are trained to identify patterns in unlabelled data, with clustering being a primary goal. This involves discovering inherent groupings within data to determine whether a data point is part of a cluster. Common examples of unsupervised algorithms include k-means clustering and principal component analysis

On the other hand, supervised learning involves training models on data that is labelled with specific outputs. Examples of supervised learning algorithms encompass K-Nearest Neighbours (KNN), Support Vector, Gradient Boosting (GB), Random Forest (RF), Linear Regression, Logistic Regression (LR), Decision Trees (DT), Neural Networks (NNs), and Naïve Bayes, each correlating input data with predetermined outcomes. The KNN model has been employed by Yuzhen et el. [37] to detect the presence, count, and position of occupants utilizing motion sensors.

The Convolutional Deep Bi-directional Long Short-Term Memory (CDBLSTM) method utilizes a convolutional neural network combined with a deep bi-directional long short-term memory network to effectively learn and encode sequential features from environmental sensor data for accurate building occupancy estimation [38]. Z. Chen et el. [39] focuses on building occupancy estimation using environmental sensors, which is crucial for enhancing energy efficiency in buildings. The research is directed at leveraging non-intrusive, low-cost sensor data to estimate occupancy levels, addressing the need for reliable occupancy information in energy-related building control systems. The methodology involves the development and application of a CDBLSTM approach. This method integrates a convolutional network for initial feature extraction and a deep structure for automatic learning of significant features from sensor data without manual intervention. Additionally, it employs Long Short-Term Memory (LSTM) networks to capture temporal dependencies in data, with a bi-directional structure considering both past and future contexts for occupancy identification. To evaluate the effectiveness of the CDBLSTM approach, the study compares it with several state-of-the-art methods, including Hidden Markov Model (HMM) with information gain-based feature selection, Decision Tree with raw data features, Elaboration likelihood model (ELM) with wrapper-based feature selection, and Linear Discriminant Analysis with raw data features. The performance of each model is assessed based on estimation accuracy, Normalized Root Mean Square Error (NRMSE), and the detection accuracy of presence/absence. The occupancy data was categorized into ranges (zero, low, medium, high) based on historical occupancy patterns. The dataset comprised 31 workdays, with the first 26 days used for training and the remaining 5 days for testing.

The research [40] focuses on eXtreme Gradient Boosting (XGBoost), a scalable tree boosting system, widely recognized for its efficiency and performance across various machine learning challenges. The system is designed to handle large-scale data efficiently and employing methodologies like regularized learning, gradient boosting, shrinkage, column subsampling, and sparsity-aware algorithms, making it a popular choice among data scientists for tasks involving complex data dependencies.

This study utilizes the efficiency and adaptability of XGBoost, assessing its performance by comparing it with various classifiers including Logistic Regression , Linear Discriminant Analysis (LDA), Classification And Regression

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Tree (CARTree), K-Nearest Neighbours , SV. The comparison focuses on evaluating each model's effectiveness using metrics such, confusion matrix, Accuracy, Precision and F1 score.

3.2 Occupancy profiling approach

In the pursuit of energy efficiency and enhanced indoor comfort, recent discussions in building automation have centred on occupant-centred control models. These models, incorporating innovative technologies, aim to dynamically regulate building systems based on precise occupant information. This review delves into the evolving landscape of occupancy profiling.

In particular, the integration of environmental sensors and machine learning algorithms has emerged as a promising method for identifying both the number of occupants and the intensity of their activities.

This exploration into occupant-centred control strategies, environmental sensing, and intelligent algorithms signifies a significant step forward in optimizing building service systems. The advancements discussed, hold promise for achieving energy-efficient, sustainable, and comfortable environments in large spaces, providing a foundation for future developments in smart building technologies [41].

The Ignacio Benítez [42] , explores the application of dynamic clustering to analyze daily load profiles of Spanish residential energy consumers over 2008 and 2009. The study utilizes dynamic clustering algorithms for classification of time-series data of load profiles, demonstrating their effectiveness as a tool for quickly categorizing clients based on energy consumption patterns and overall trends. The Prabhakaran's research [43] focuses on applying these techniques to the task of occupancy estimation in buildings, which is a crucial component in energy management systems. The authors propose a novel approach that utilizes an unsupervised and interpretable machine learning method, specifically the Explainable K-Means Clustering (ExKMC) algorithm, for estimating occupancy in buildings. Traditional clustering methods can often yield results that are difficult to interpret or rely on data that are hard to label. To address these challenges, the researchers use a small binary decision tree for clustering, which improves the interpretability of the results. The ExKMC algorithm works by creating a small tree with a predefined number of leaves (k) that partitions the data into clusters. This tree is then expanded to produce a new tree with more leaves (k'), resulting in clusters that are both accurate and explainable. The methodology involves a trade-off between prediction accuracy and the ease of interpreting the clustering decisions.

This his research, Vassiljeva [44] focuses on developing an occupancy-based algorithm for optimizing ventilation systems in buildings, specifically in a school environment in Estonia. The aim is to reduce energy consumption in line with the European Green Deal targets by 2050 while maintaining indoor comfort. The methodology involves several steps:

1) Data Collection and Cleaning: Electricity smart meters provide detailed hourly power consumption data from various points, including the main meter and ventilation units. This data is cleaned to eliminate inconsistencies.
2) Clustering for Consumption Profiles: Using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm, electricity consumption data is clustered to form distinct profiles for different times (workdays, weekends, vacations).
3) Occupancy Detection and Estimation: The research develops occupancy profiles based on electricity consumption, using a threshold method for detection and a proportional method for estimation.
4) Developing Occupancy Schedules: These profiles are used to create schedules indicating expected occupancy.
5) Ventilation Schedule Formulation: The study proposes two ventilation strategies - Classical and Demand-Controlled Ventilation (DCV). The Classical approach uses an on-off method based on occupancy detection, while DCV adjusts airflow rates based on occupancy estimation, considering factors like minimum airflow and air quality standards.

This research presents a data-driven approach to optimize building ventilation, offering a method adaptable to various building types and aligning with sustainable operation goals.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

The DBSCAN algorithm will be utilized in the current implementation in terms of T4.2, to cluster the occupancy level in a building, as outlined in the aforementioned study. The electricity consumption and CO2 concentration data will be categorized into distinct profiles using the DBSCAN clustering algorithm. These profiles will reflect the varying occupancy patterns within the building for different periods. The clustered data will then be employed to create occupancy profiles, which will be developed through methods such as threshold-based detection and ratio-based estimation, providing insights into the occupancy levels at different times. This approach is aligned with contemporary energy conservation and sustainability goals, making it a relevant and valuable contribution to efficient building management.

4 SmartLivingEPC CIEM

Common Information Exchange Model (CIEM) constitutes a cloud-based middleware with semantic capabilities, encompassing a wide range of services for data ingestion, management, and sharing, serving as the backbone for seamless data exchange within the SmartLivingEPC project. In the sections below, an overview of the six-layered CIEM architecture is provided. Furthermore, the interfaces that define the high-level intra- and extra-component interactions are outlined. Note that, intra-component interactions refer to the interactions that take place within the CIEM and involve the different CIEM sub-components, as these are described further on in the deliverable. Similarly, the extra-component interactions, refer to the interactions (data requests, data exchanges etc.) that occur between CIEM and the various SmartLivingEPC components.

4.1 CIEM Data model

In this section, a high-level description of the CIEM data model to be used within the framework of the SmartLivingEPC project is presented. Based on the literature review of the available BIM-IoT integration methods, addressed in Figure 5, the proposed methodology is based on the BOT and Brick ontology to provide granular representation of building systems while enabling seamless integration and analysis across the smart infrastructure deployed at the SmartLivingEPC pilots.

Figure 5 displays the main elements of the current state of the CIEM data model, as well as the relationships between them. The presented data model is developed to perform as the core module of the SmartLivingEPC CIEM ontology. The proposed data model, leveraging the BOT ontology describes the topology of a building as well as the relationships between their main components i.e., site, building, storey and space. In the proposed CIEM Data Model, a “Prosumer”-centric approach is followed, facilitating the interconnection between the building topology and the IoT infrastructure.

On the other hand, the Brick schema is chosen to describe the various components and aspects of the smart buildings and their operations. Leveraging the Brick schema, different equipment, sensors, actuators, and systems within a building can be defined and understood in a standardized way. This structured framework shall facilitate the integration and management of the diverse range of devices and data sources present in the IoT-enabled SmartLivingEPC buildings, making it easier to develop applications for asset and operational rating, energy management, and overall optimization.

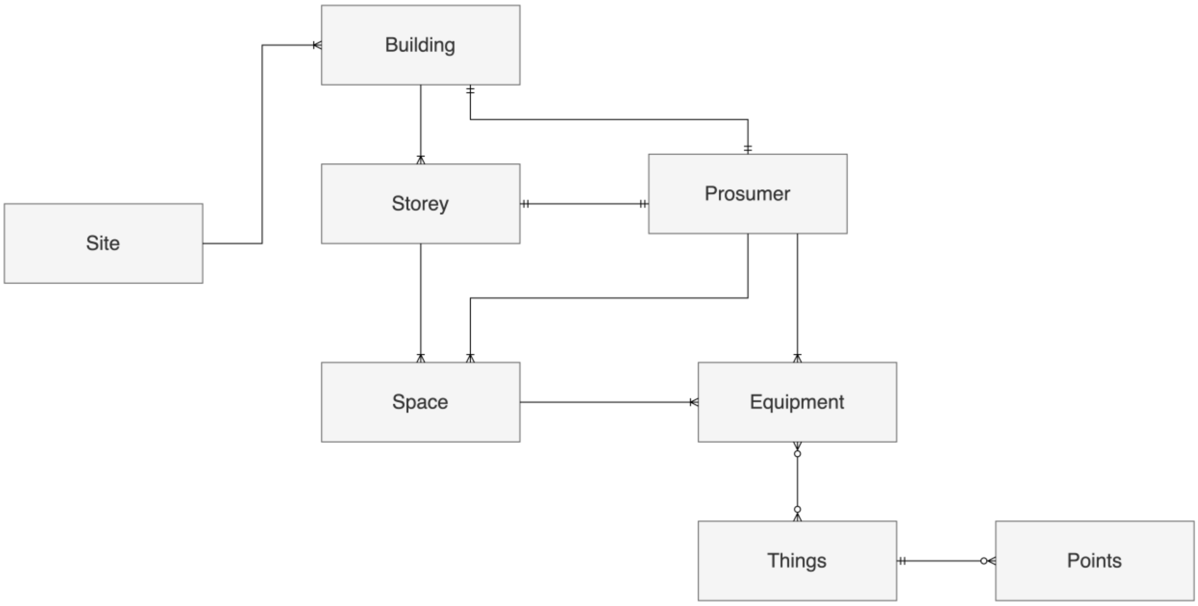

Figure 5: High-level description of the proposed CIEM data model

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

4.2 CIEM Data management

The SmartLivingEPC CIEM is a cloud-based solution that concentrates a collection of services required to collect, manage and share the static and dynamic data, while ultimately serving the SmartLivingEPC architecture. In this section, a high-level description of the multi-layered architecture is presented along with the specifications related to the implementation of CIEM within the SmartLivingEPC framework.

4.2.1 Multi-layered architecture

CIEM is cloud-based solution where project information (e.g., energy consumption, generation and storage, non-energy sources consumption), documents (e.g. audits), and data (BIM and IoT) are stored, managed, and exchanged by all stakeholders throughout the project lifecycle. It aims at facilitating collaboration, coordination, and information sharing among the various parties involved. To seamlessly perform its activities and serve the SmartLivingEPC architecture, CIEM is structured based on a multi-layered architecture. Figure 6 presents an overview of the high-level multi-layered architecture of the CIEM, including CIEM's sub-components (hereafter referred to as layers) and interactions with the rest of the SmartLivingEPC components.

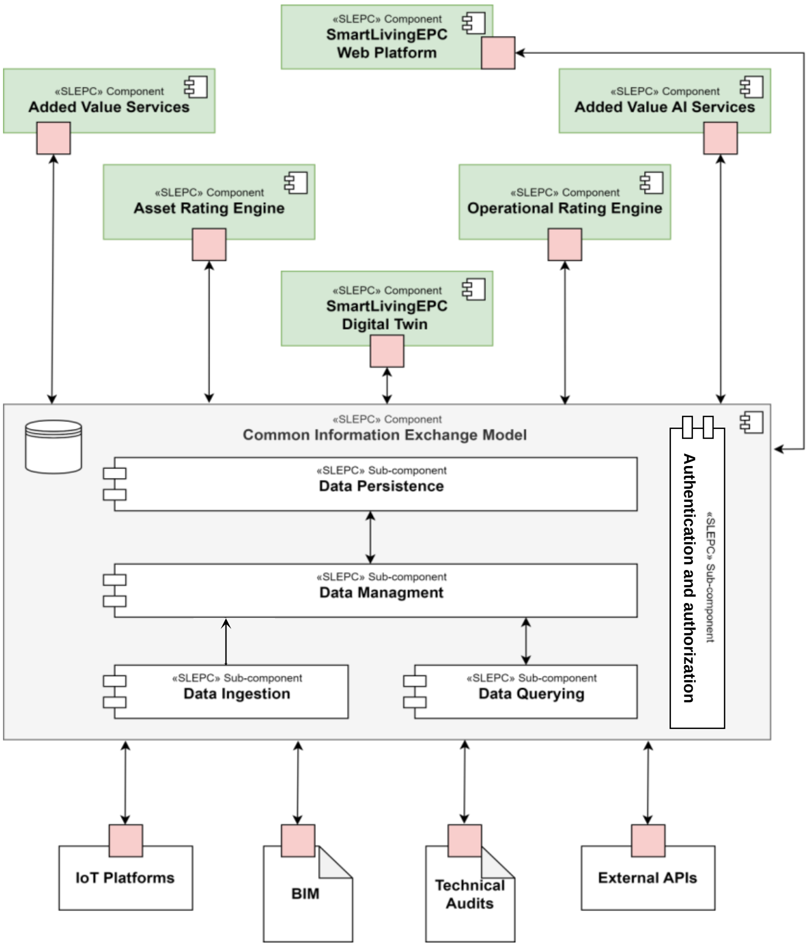

Figure 6: SmartLivingEPC CIEM functional diagram

Four layers are identified in the first version of the CIEM namely, the Data Ingestion Layer, Data Querying Layer, Data Management Layer and Data Persistence Layer. These layers are briefly discussed here:

- Authentication and Authorization Layer: The Authentication and Authorisation Layer oversees the storage and administration of user accounts and their roles within the SmartLivingEPC framework, facilitating user registration and verification by the CIEM.
- Data Ingestion Layer: All (static and dynamic) data from various sources (IoT platforms and external APIs) is ingested before it can be processed and stored. CIEM's Data Ingestion layer handles the aforementioned process. Note that Data Ingestion is responsible for the data that is sent to the CIEM without any prior query/request from the latter. All the ingested data must be in line with the specified SmartLivingEPC data model. This layer interfaces internally with the Data Management layer and externally with services and applications that are outside the CIEM component;
- Data Querying Layer: As opposed to Data Ingestion, the Data Querying layer of CIEM is responsible for querying data from external sources in order to cover the needs of the various SmartLivingEPC components. Also, in this case, the ingested data must be in line with the specified SmartLivingEPC data model. Data Querying is integrated with services and applications outside the CIEM component besides the internal communication with the Data Management layer;
- Data Management Layer: Within the Data Management layer, processing of the ingested data is performed. Data processing includes, among others, data cleansing, data normalising, parsing of IFC files, performing consistency, completeness, and correctness checks of the IFC files etc. Note that this layer interfaces only with the Data Persistence layer;
- Data Persistence Layer: This layer consists of various types of datastores, each pertaining to specific groups of (structured) data that are essential to cover the needs of the SmartLivingEPC applications and services. Examples of such data storage are file storage databases, time-series databases (handling relevant data generated by the IoT services) and key-value databases.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

4.2.2 High-level description of the fused BIM-IoT data management

As noted previously, CIEM will be based on the fused management of BIM and IoT data. Hereafter a systematic approach is explored in order to provide an efficient representation of a framework that efficiently receives, processes and translates static and dynamic data into an understandable form by other SmartLivingEPC components.

In Figure 7, an overview of the SmartLivingEPC BIM data management framework is displayed. The proposed software implementation leverages RESTful APIs for retrieving IFC files. Prior to the retrieval, the Authentication and Authorisation layer ensures that only legitimate users or systems with valid credentials can access the exposed APIs, preventing thus unauthorized access and protecting sensitive BIM data. Once retrieved and authenticated, BIM data undergoes compliance checking. Within the context of BIM and IFC, compliance checking involves the thorough assessment of a building model to ensure it aligns with predefined rules, standards, or regulations aiming at assessment of a building design with respect to the configuration of objects, their respective relations or attributes [45]. This process scrutinizes elements within the model, including architectural, structural, and system components, to verify their compliance with specific criteria such as spatial relationships, material specifications, and energy efficiency. The aim is to guarantee that the building design adheres to the necessary requirements set by the IFC schema. IFC, serving as an interoperable data exchange format in BIM, facilitates this evaluation by employing the IfcOpenShell library (as also done in the study by [46]) and in the COGITO project, for serialising/deserialising the IFC data as well as for the assessment of the model's STEP data compliance against the established IFC4 EXPRESS schema. The assessment involves checking the related datatypes, class names, range of numerical variables and the sizes of the data collections. Following the Compliance Checker, the IFC file is afterwards checked in order to verify its completeness against predefined concepts. The BIM data management employs the relevant data repository to store temporary files and the key-value database to store the IFC objects.

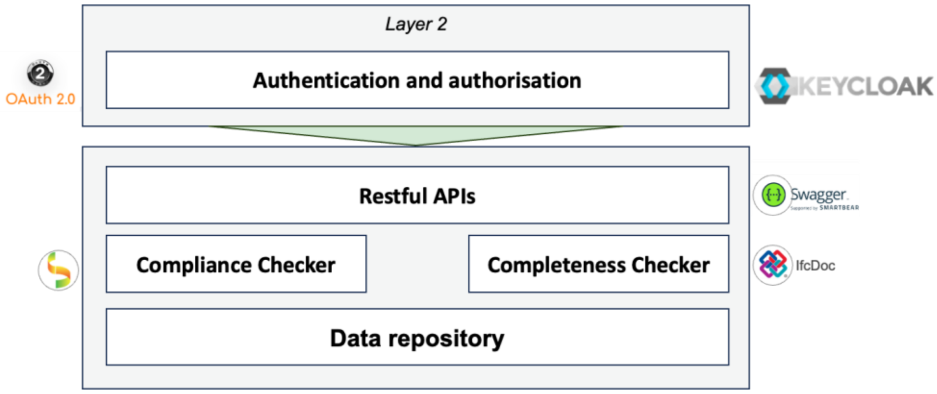

Figure 7: SmartLivingEPC CIEM BIM data management

The IoT data collection process is initiated via a message broker that communicates with the various SmartLivingEPC IoT data sources presented in D1.2 (as part of the Data Collection Layer). The aforementioned sources are integrated with the message broker using specific protocols that are described in Section 4.2.3. Acting as an intermediary, the message broker receives (Layer 1) the queued data. Once gathered, this influx of data moves through the Authentication and Authorization Layer (Layer 2), ensuring that only authenticated and authorized data gains access to the storage and management systems (Layer 3), verifying the legitimacy of incoming data streams, enforcing security protocols, and granting access rights based on predefined permissions. Once verified, data is then routed to designated databases, based on predefined rules or topics, where it is stored. Initial processing tasks like data filtering, normalization, or aggregation are here foreseen in order to make the ingested data more usable and efficient for further analysis. The aim is to further refine raw IoT data, paving the way for efficient storage, analysis, and visualization by enhancing data quality, consistency, and relevance. Effective initial processing ensures that downstream analytics and decision-making processes are based on accurate, meaningful, and actionable information extracted from IoT data. Once IoT data is processed, it is then rerouted to the other SmartLivingEPC components/systems through RESTful APIs, establishing a standardized communication method and seamless data exchange. Eventually, the processed data, is sent/retrieved through designated API endpoints, allowing for tailored access to specific datasets. This approach ensures a streamlined and standardized way for different software components to interact with the refined IoT data, facilitating its utilization across the SmartLivingEPC system.

An overview of the abovementioned data flow is depicted in Figure 8.

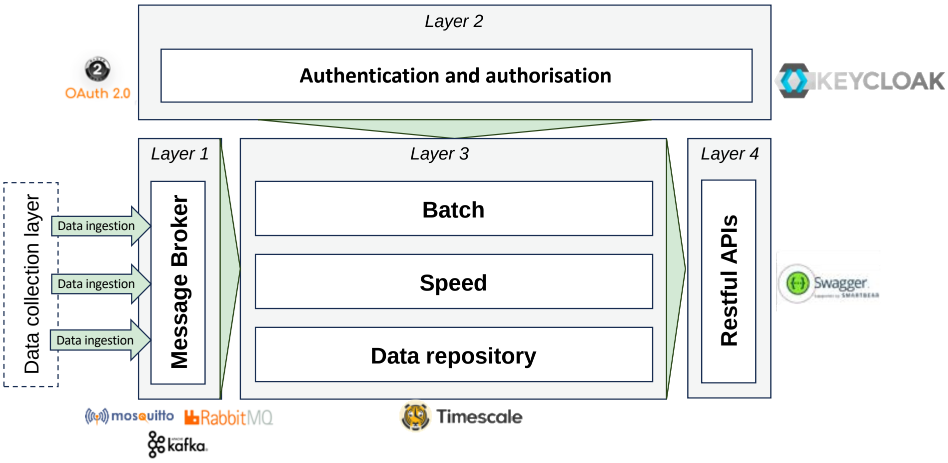

Figure 8: SmartLivingEPC CIEM IoT data management

Eventually, the fused BIM-IoT data management system will integrate the capabilities of the above-described BIM and IoT data management workflows in order to a create a comprehensive solution for managing, primarily, the operation of the SmartLivingEPC pilot buildings. The fused BIM-IoT data management, depicted in Figure 9, shall enable simultaneous management of both BIM and IoT data. Ultimately, the combined analysis will ensure that the BIM model reflects the most up-to-date information about the building's status and performance, by seamlessly interelating the IoT with the BIM data (supported as well by the CIEM data model).

The simultaneous management of BIM and IoT data in a fused system is expected to enhance the effectiveness of decision-making, facilitate proactive maintenance, optimize building operations, and foster a more holistic understanding of the building's lifecycle.

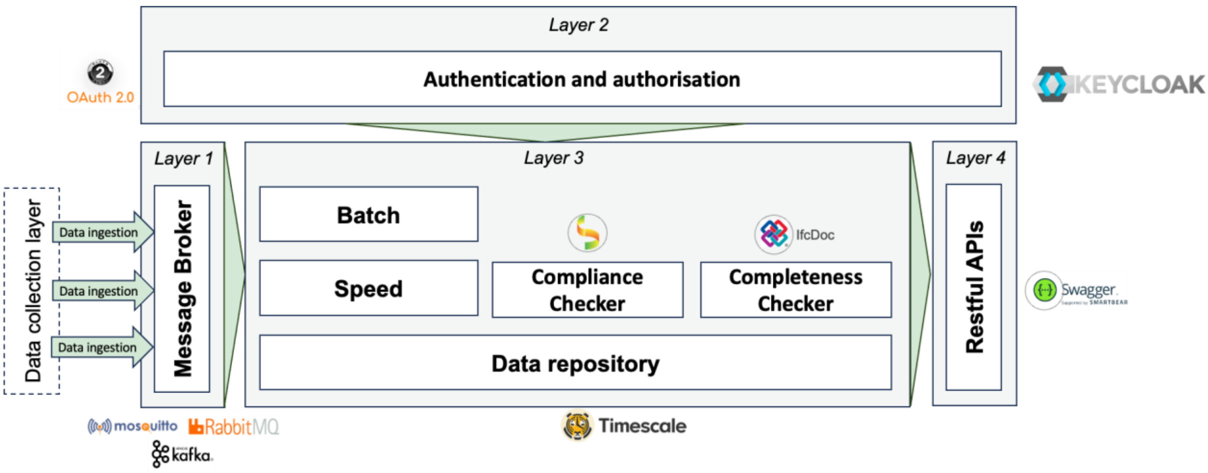

Figure 9: SmartLivingEPC CIEM fused BIM-IoT data management

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

4.2.3 Technology stack and implementation

CIEM platform builds upon a common data environment that QUE has developed, delivered and deployed initially for IoT static and dynamic data management, following the Lambda Architecture and demonstrated in the context of the COGITO project[20]. The data from the various sources is ingested into a queue of messages to be processed before being permanently stored within the CIEM repository. The IoT sensorial middleware employs, for the asynchronous communication, message brokers such as RabbitMQ, Mosquitto MQTT and Apache Kafka. The Data Processing Layer described earlier consists of two sub-layers; the batch and the speed layer (see Figure 8). In the Batch layer, data stored in the data repository or directly received from the message brokers are divided into small batches and sent for processing. The aforementioned sub-layer mainly targets the processing of large datasets and production of detailed outputs. Once a batch is processed, batch results are created and pushed to the Service layer or stored in the CIEM repositories. The Speed layer, responsible for all the lightweight (near) real-time data-processing (e.g., cleansing, normalising), builds upon state-of-the-art technologies namely Apache Spark and Kafka. To exchange historical data with the various SmartLivingEPC components, CIEM exposes a unified set of APIs via API query languages. To this end, the Swagger Interface Description Language is used as a means of automated API documentation and automated code generation into any programming language.

To support the management of BIM data, CIEM utilizes the open-source software library IfcOpenShell[21] that enables the reading, writing, and manipulation of IFC files. IfcOpenShell is currently available on GitHub[22]. It provides a set of tools and functions that facilitate the handling of IFC files, making it easier to work with BIM data across various software applications and platforms. IfcOpenShell is available in Python. In this case, Python v3.7 is employed. To define the Model View Definitions (MVD), necessary for the completeness checking of the IFC files, the IfcDoc tool is considered.

Within the Authentication and Authorization Layer of CIEM, Keycloak[23] serves as a fundamental component within an Authentication and Authorization Infrastructure (AAI) ensuring a robust data management platform. With its open-source architecture, Keycloak provides a comprehensive suite of identity and access management tools, enabling seamless authentication and fine-grained authorization controls across diverse applications and services. As an integral part of the AAI, Keycloak facilitates centralized user management, implements secure single sign-on (SSO), supports integration with various identity providers, and ensures stringent access control mechanisms. Its versatility and scalability make it an essential layer within the CIEM, empowering its security while simplifying user access to data and resources across the ecosystem.

An overview of the technologies/libraries used as part of the CIEM development is listed in Table 1.

Table 1: SmartLivingEPC CIEM technologies and libraries

Technology/library	Version	License
Apache Tomcat	latest	Apache License 2.0 license
Apache 2 (HTTP server)	latest	Apache License 2.0 license
Apache Spark	latest	Apache License 2.0 license
Apache Kafka	latest	Apache License 2.0 license

[20] https://cogito-project.eu/

[21] https://ifcopenshell.org/

[22] https://github.com/IfcOpenShell/IfcOpenShell

[23] https://www.keycloak.org

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Oracle Java	latest	GPLv2 (GNU General Public License)
Spring MVC	latest	Apache License 2.0 license
PostgreSQL	latest	PostgreSQL license (similar to BSD/MIT)
Hibernate	latest	LGPL-2.1 license
RabbitMQ	latest	MPL 2.0 license
MQTT	latest	GPLv3 (GNU General Public License)
NGiNX	latest	2-clause BSD license
IfcOpenShell	0.7.0	LGPL-3.0 License

4.2.4 Stakeholder roles

Definition of stakeholder roles is based on the outcomes of Task 1.2 "*Elicitation of stakeholders requirements & market needs*" reported within D1.1 "*From EPC schemes gaps and opportunities to SmartLivingEPC requirements, recommendations and market needs*". The stakeholders involved represent the end-users of the various SmartLivingEPC applications/tools and main actors under the SmartLivingEPC framework. Thus, the main stakeholder roles are the following: Building owners/tenants, authorities, building unit assessors, building complex assessors.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

4.2.5 CIEM interconnection with pilot sites

As presented in D1.2, SmartLiving framework will be evaluated on three building units (i.e., Pilot #1 nZEB Smart House DIH, Pilot #2 Frederick's University Main Building, Pilot #3 Ehituse Maemaja, Tallinn University of Technology) and one building complex (i.e., Building Complex of Leitza). The initial interconnection was established with Pilot #1 nZEB Smart House DIH.

4.2.5.1 BIM/ IFC files

As previously stated in Figure 10, the IFC files need to be obtained from the pilots. Subsequently, after verifying their content, these files should be converted into an understandable format (aligned with the CIEM's data model) and then made available to the SmartLivingEPC components. As far as BIM files are concerned and within the frames of SmartLivingEPC, the IFC4 ADD2 TC1 schema is employed. IFC4 ADD2 TC1 represents the latest official schema based on the ISO 16739-1:2018[24].

An example of an IFC file, representative of the Greek pilot site situated in Thessaloniki, which is the main pilot site of SmartLiving EPC project, namely the nZEB Smart House DIH is shown in Figure 10.

```
ISO-10303-21;
HEADER;
FILE_DESCRIPTION(('ViewDefinition [DesignTransferView_V1.0]','RevitIdentifiers [VersionGUID: 2bf03851-d2e8-4f9e-9b2a-eb5458631717, NumberOfSaves: 71]','CoordinateReference [CoordinateBase: Shared Coordinates, ProjectSite: Internal]'),'2;1');
FILE_NAME('Project Number','2023-04-24T14:32:45+02:00',(''),(''),'ODA SDAI 22.12','23.1.10.4 - Exporter 23.2.3.0 - Alternate UI 23.2.3.0','');
FILE_SCHEMA(('IFC4'));
ENDSEC;

DATA;
#1=IFCORGANIZATION($,'Autodesk Revit 2023 (ENU)',$,$,$);
#2=IFCAPPLICATION(#1,'2023','Autodesk Revit 2023 (ENU)','Revit');
#3=IFCCARTESIANPOINT((0.,0.,0.));
#4=IFCCARTESIANPOINT((0.,0.));
#5=IFCDIRECTION((1.,0.,0.));
#6=IFCDIRECTION((-1.,0.,0.));
#7=IFCDIRECTION((0.,1.,0.));
#8=IFCDIRECTION((0.,-1.,0.));
#9=IFCDIRECTION((0.,0.,1.));
#10=IFCDIRECTION((0.,0.,-1.));
#11=IFCDIRECTION((1.,0.));
#12=IFCDIRECTION((-1.,0.));
#13=IFCDIRECTION((0.,1.));
#14=IFCDIRECTION((0.,-1.));
#15=IFCPERSON($,'','georgia.papadop.98',$,$,$,$,$);
#16=IFCORGANIZATION($,'','',$,$);
#17=IFCPERSONANDORGANIZATION(#15,#16,$);
#18=IFCOWNERHISTORY(#17,#2,$,.NOCHANGE.,$,$,$,1682324390);
#19=IFCSIUNIT(*,.LENGTHUNIT.,$,.METRE.);
#20=IFCSIUNIT(*,.AREAUNIT.,$,.SQUARE_METRE.);
#21=IFCSIUNIT(*,.VOLUMEUNIT.,$,.CUBIC_METRE.);
#22=IFCSIUNIT(*,.PLANEANGLEUNIT.,$,.RADIAN.);
#23=IFCDIMENSIONALEXPONENTS(0,0,0,0,0,0,0);
#24=IFCMEASUREWITHUNIT(IFCRATIOMEASURE(0.0174532925199433278),#22);
#25=IFCCONVERSIONBASEDUNIT(#23,.PLANEANGLEUNIT.,'DEGREE',#24);
#26=IFCSIUNIT(*,.MASSUNIT.,.KILO.,.GRAM.);
#27=IFCDERIVEDUNITELEMENT(#26,1);
#28=IFCDERIVEDUNITELEMENT(#19,-3);
#29=IFCDERIVEDUNIT((#27,#28),.MASSDENSITYUNIT.,$);
#30=IFCDERIVEDUNIT((#27,#28),.IONCONCENTRATIONUNIT.,$);
#31=IFCDERIVEDUNITELEMENT(#19,4);
#32=IFCDERIVEDUNIT((#31),.MOMENTOFINERTIAUNIT.,$);
#33=IFCSIUNIT(*,.TIMEUNIT.,$,.SECOND.);
#34=IFCSIUNIT(*,.FREQUENCYUNIT.,$,.HERTZ.);
#35=IFCSIUNIT(*,.THERMODYNAMICTEMPERATUREUNIT.,$,.KELVIN.);
#36=IFCSIUNIT(*,.THERMODYNAMICTEMPERATUREUNIT.,$,.DEGREE_CELSIUS.);
#37=IFCDERIVEDUNITELEMENT(#35,-1);
#38=IFCDERIVEDUNITELEMENT(#33,-3);
#39=IFCDERIVEDUNIT((#27,#37,#38),.THERMALTRANSMITTANCEUNIT.,$);
#40=IFCDERIVEDUNITELEMENT(#19,1);
#41=IFCDERIVEDUNIT((#27,#37,#38,#40),.THERMALCONDUCTANCEUNIT.,$);
#42=IFCSIUNIT(*,.LENGTHUNIT.,.DECI.,.METRE.);
#43=IFCDERIVEDUNITELEMENT(#19,3);
#44=IFCDERIVEDUNITELEMENT(#33,-1);
#45=IFCDERIVEDUNIT((#43,#44),.VOLUMETRICFLOWRATEUNIT.,$);
#46=IFCDERIVEDUNIT((#27,#44),.MASSFLOWRATEUNIT.,$);
#47=IFCDERIVEDUNIT((#44),.ROTATIONALFREQUENCYUNIT.,$);
#48=IFCSIUNIT(*,.ELECTRICCURRENTUNIT.,$,.AMPERE.);
```

Figure 10: IFC representation of the Greek pilot site

[24] https://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications/

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

4.2.5.2 IoT infrastructure

As noted in Section 4.2.3, the communication between CIEM and IoT platforms should be based on asynchronous event-driven communication techniques and based on a MQTT broker for real-time data exchange, while historical data will be exchanged through RESTful services. Requirements concerning the intercommunication between CIEM and the IoT infrastructures have already been shared by QUE to the respective pilot partners. Due to reasons of brevity here, only the most important information is presented.

Service intercommunication consists of three major categories:

1. **Static information**. Information related to the fixed, unchanging data or details associated with the IoT infrastructure *(From third-party service to CIEM)*;

An example of static configuration is displayed in Figure 11 as this is retrieved via RESTful API from the Greek pilot's IoT infrastructure.


```json
[{
        "building": "134a31dVLAMBYUShIBlUJY",
        "device": {
            "id": "d381989004dc45658d359896532a41c9",
            "name": "MCOHome Multisensor",
            "public_id": "7f3da9a1-cfaf-472e-b703-8a81775a5ac1",
            "type": ["CO2SENSOR", "HUMIDITYSENSOR", "PM25SENSOR", "TVOCSSENSOR", "LIGHTSENSOR", "MOVEMENTSENSOR", "SOUNDSENSOR", "TEMPERATURESENSOR", "SMOKESENSOR"]
        },
        "space": {
            "id": "1F8oBIUbz1LwYG6clvlyFY",
            "name": "living_room_26"
        },
        "zone": {
            "id": "2XrgBF$IfB8PcTbtAjULke",
            "name": "Residential"
        }
    }, {
        "building": "134a31dVLAMBYUShIBlUJY",
        "device": {
            "id": "1F8oBIUbz1LwYG6clvlyEz",
            "name": "Electrical_Controls_Fibaro_Motion-sensor.eng:Default:1017561",
            "public_id": "ec89b4ddbdc74e289585e06f274cc033:1F8oBIUbz1LwYG6clvlyEz",
            "type": ["MOVEMENTSENSOR", "TEMPERATURESENSOR", "LIGHTSENSOR"]
        },
        "space": {
            "id": "1F8oBIUbz1LwYG6clvlyFY",
            "name": "living_room_26"
        },
        "zone": {
            "id": "2XrgBF$IfB8PcTbtAjULke",
            "name": "Residential"
        }
    }, {
        "building": "134a31dVLAMBYUShIBlUJY",
        "device": {
            "id": "1F8oBIUbz1LwYG6clvlyEu",
            "name": "Magnetic contact FIBARO v2 with extrusion:SensorDoorWindoeRF433/88190WSmartHomeWi-Fi:1017564",
            "public_id": "ec89b4ddbdc74e289585e06f274cc033:1F8oBIUbz1LwYG6clvlyEu",
            "type": ["CONTACTSENSOR"]
        },
        "space": {
            "id": "1F8oBIUbz1LwYG6clvlyFY",
            "name": "living_room_26"
        },
        "zone": {
            "id": "2XrgBF$IfB8PcTbtAjULke",
            "name": "Residential"
        }
    }, {
        "building": "134a31dVLAMBYUShIBlUJY",
        "device": {
            "id": "2w48yaP$r6tRw2QCjYS_GO",
            "name": "CO2 Sensor:SH_C02_Temp:1026738",
            "public_id": "0916e774-f7cc-4c4a-bf26-0cf5ed782a24",
            "type": ["TEMPERATURESENSOR", "CO2SENSOR"]
        },
        "space": {
            "id": "1F8oBIUbz1LwYG6clvlyFY",
            "name": "living_room_26"
        },
        "zone": {
            "id": "2XrgBF$IfB8PcTbtAjULke",
            "name": "Residential"
        }
```


Figure 11: Part of the Greek pilot's IoT infrastructure static configuration

2. **Data Acquisition**. In this case, the connected devices are publishing real-time values on-change or historical time-series data *(From third-party service to CIEM)*;

For the communication related to time-series data, direct exchange or topic exchange is used[25]. A direct exchange delivers messages to queues based on the message routing key and is ideal for the unicast routing of

[25] https://www.rabbitmq.com/getstarted.html

messages (although they can be used for multicast routing as well). Furthermore, they are often used to distribute tasks between multiple workers (instances of the same application) in a round-robin manner.

On the other hand, topic exchanges route messages to one or many queues based on matching between a message routing key and the pattern that was used to bind a queue to an exchange. The topic exchange type is often used to implement various publish/subscribe pattern variations and are commonly used for the multicast routing of messages.

Topic exchanges have a very broad set of use cases. Whenever a problem involves multiple consumers/applications that selectively choose which type of messages they want to receive, the use of topic exchanges is considered.

Specifically, every new data produced by each one of the available metrics the connected devices have is then published to a predefined queue using the following message structure in JSON format:

Table 2: Time-series message structure

Time-series data communication message	
item	String
source	String
value	String
timestamp	String

Assuming an ambient sensing device capable of measuring temperature, relative humidity, luminance and occupancy, configured to publish all its measurements as configured. Every time the device publishes its measurements, the following four messages are published in the respective queue:

```
{
  "item": "sensorTemperature_ID",
  "source": "device",
  "value": "27.5",
  "timestamp": "2021-05-25T13:33:33.000Z"
}
```

Figure 12: Sample message for temperature measurement

```
{
  "item": "sensorRelHumidity_ID",
  "source": "device",
  "value": "50",
  "timestamp": "2021-05-25T13:33:33.000Z"
}
```

Figure 13: Sample message for relative humidity measurement

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

```
{
  "item": "sensorLuminance_ID",
  "source": "device",
  "value": "254",
  "timestamp": "2021-05-25T13:33:33.000Z"
}
```


Figure 14: Sample message for luminance measurement


```
{
  "item": "motionAlarm_ID",
  "source": "device",
  "value": "1",
  "timestamp": "2021-05-25T13:33:33.000Z"
}
```


Figure 15: Sample message for occupancy measurement

3. **Acquisition of connection status updates**. *(From third-party service to CIEM)*.

For this communication category, direct exchange is used. Specifically, every time the connection status of any connected device changes, a message is published to a predefined queue highlighting its current connection status in JSON format as follows:

Table 3: Connection status message structure

Connection status message	
Serial	**String**
thingUID	**String**
Status	**String (*allowable values: [ONLINE, OFFLINE]*)**
Description	**String**
Timestamp	**String**

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

5 Building Dynamic Behaviour Monitoring System: Data acquisition

The development of a Building Dynamic Behaviour Monitoring System closely integrates the gathering and examination of multiple environmental factors with sophisticated data analytics techniques. The goal is to decode the intricate relationships between occupancy trends, energy usage, and the quality of the indoor environment.

At the forefront of our investigation lies the process of preliminary testing. This critical phase brings the transition from raw data acquisition to the development of predictive models. Through a methodical examination of diverse attributes – including temperature, relative humidity, light levels and CO2 concentrations – this research investigates the data integrity, feature relevance, and temporal dynamics. These attributes extend well beyond basic data points since they are the essential elements that construct the framework of our understanding about room occupancy and its relationship with the building's environmental and energy characteristics.

The nZEB Smart House, as main demonstration site, used to evaluate a variety of metrics, from HVAC system loads to ambient environmental factors, and overall energy usage. The dataset's temporal granularity was adjusted, spanning from event-based records to minute-by-minute logs. This diverse range of data collection ensures a thorough representation of the smart house's dynamic environment.

5.1 UCI repository

In the early stages of our project, we engaged in preliminary testing [47]—a critical phase that bridges the gap between raw data acquisition and advanced model training. Preliminary testing involves an evaluation of the acquired data, ensuring its suitability for the intended binary classification task.

During the preliminary testing phase, each attribute is carefully analysed to determine its influence on room occupancy prediction. The attributes that analysed are:

Timestamps: The inclusion of detailed timestamps is indispensable for recognizing usage patterns and assessing how other environmental variables change over time relative to occupancy.

Temperature: Since human presence often correlates with variations in temperature, this attribute can be a strong predictor of occupancy when analysed over time.

Relative Humidity: Humidity levels can fluctuate with occupancy due to human respiration and perspiration, making it a valuable feature in our dataset.

Light: Light levels can directly indicate occupancy, especially in contexts where human activity is linked to light usage.

CO2 Levels: CO2 concentration is a direct biomarker of human presence. By monitoring changes in CO2 levels, we can infer occupancy with high confidence.

Humidity Ratio: Variations in the humidity ratio may also reflect occupancy, as it integrates both temperature and relative humidity, which are affected by human activity.

Preliminary testing serves several essential functions in our project's lifecycle. Specifically:

1. **Data Integrity Check**: It enables us to verify the integrity and quality of the data, ensuring that all readings are consistent and accurately recorded. This step is crucial in avoiding the 'garbage in, garbage out' pitfall.
2. **Feature Relevance Assessment**: By examining each attribute, such as temperature, light, and CO2 levels, we can ascertain their relevance and predictive power concerning the occupancy status.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

3. **Temporal Dynamics Analysis**: The timestamped nature of the dataset allows us to understand the occupancy patterns over time, a vital component in creating a robust predictive model.

4. **Model Readiness**: Preliminary testing helps determine whether the dataset can be effectively split into training and testing subsets, a process that underpins the model's evaluation.

5. **Implementing Preliminary Testing**: During this phase, we have implemented the following processes:

• **Data Preprocessing**: This step includes normalizing and scaling the variables, handling missing values, and encoding categorical data, preparing the dataset for machine learning algorithms.

• **Training/Testing Split**: We have deliberately partitioned the data into a larger training set and a smaller testing set, with 8,143 and 2,664 time series instances, respectively. This split is designed to optimize the learning process while retaining a substantive portion of data for unbiased evaluation.

The preliminary testing has not only provided our team with a deep understanding of the dataset's characteristics but has also laid the groundwork for subsequent modelling phases. We have leveraged this stage to fine-tune our preprocessing pipelines and to set benchmarks for model performance. Moreover, it has facilitated a smoother transition into the iterative cycles of model training, validation, and testing, which will ultimately lead to the deployment of a reliable room occupancy classification system.

5.2 Demo Site #1: nZEB Smart House

The nZEB Smart House, as project’s primary demonstration site, stands at the forefront of our exploration into the realms of machine learning and sustainable building management.

The provision of the required information for the Building Dynamic Behaviour Monitoring Tool will be implemented by CIEM, as described in previous sections.

5.2.1 Load Category

In this pursuit, we categorize and analyse various load types within the nZEB Smart House, encompassing HVAC systems, ambient environmental parameters, and overall energy consumption. Each of these components contributes to a comprehensive understanding of the building's energy profile. The role of CIEM here is crucial. It acts as the central hub for aggregating and standardizing this diverse range of data, ensuring consistent and efficient data processing and exchange. By feeding this standardized data into our machine learning algorithms, CIEM will anable more accurate predictions and effective system optimizations. Specifically:

- **HVAC (Heating, Ventilation, and Air Conditioning)**: Capturing HVAC status, set temperatures, and fan speeds allows for the analysis of energy consumption patterns and the development of predictive models for system optimization.
- **Ambient Sensing**: Environmental parameters such as temperature, humidity, luminance, CO2 concentration, and occupancy/presence are monitored to ensure comfort and air quality while maximizing energy savings.
- **Total Metering**: The overall energy consumption measured in kilowatt-hours (kWh) provides a holistic view of the house's energy profile.

5.2.2 Temporal granularity (minimum)

The temporal granularity of our dataset is meticulously structured to capture the dynamic environment of the nZEB Smart House. Specifically:

- **Event-Based:** Certain metrics such as HVAC status, lighting status (dimming), and presence are recorded based on events, ensuring detailed tracking of changes as they occur.

The SmartLivingEPC project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101069639.

- **1-5 Minutes:** The HVAC set temperature is logged with a granularity ranging from 1 to 5 minutes, providing near real-time data for responsive system adjustments.
- **5 Minutes:** Temperature, relative humidity, and CO2 levels are recorded every 5 minutes, striking a balance between data richness and manageability.
- **5-10 Seconds**: Luminance levels are sampled every 5 to 10 seconds, reflecting rapid changes in lighting conditions and allowing for immediate adjustments in lighting control systems.
- **15 Minutes:** Energy consumption and CO2 data are collected every 15 minutes, providing a comprehensive overview of long-term trends without overwhelming the data storage and processing capabilities.

6 Methodological approach for the Building Dynamic Behaviour Monitoring Systems

The Building Dynamic Behaviour Monitoring System integrates two key components: the Analytics and Modeling sub-component, along with the Dynamic Behaviour Generation sub-component. This methodology was designed not only to comprehend intricate patterns of energy usage and occupancy but also to anticipate the building's behavior. By synthesizing data-driven insights with predictive analytics, it laid the foundation for optimizing building operations. The system designed for SmartlivingEPC project employs advanced analytics to model the present state, while the Dynamic Behaviour Generator forecasts future usage, ensuring efficiency and sustainability in building management.

In the model illustrated in Figure 16, a functional diagram for a Building Dynamic Behaviour Monitoring System is shown. The CIEM acts as the foundational gateway through which raw data inflow is channelled. The Building Dynamic Behaviour Monitoring System was intricately designed to process the data provided by the CIEM. Within this system, there are two critical sub-components: Analytics and Modelling, and Building Dynamic Behaviour Generator. The Analytics and Modelling sub-component is responsible for in-depth analysis and modeling of the building's dynamic behavior, transforming raw data into actionable insights. The Building Dynamic Behaviour Generator complements this by generating dynamic behavior based on analyzed data, providing a predictive preview into the building's operational dynamics.

Once these sub-components have effectively processed the data and synthesized the results, the outcomes are then conveyed to the SmartLivingEPC Web Platform. This platform serves as the user interface for the system, presenting processed data and insights in a user-accessible format, enabling users to interact with the data through a web-based portal. This seamless integration of data processing and user-centric display epitomizes the model's efficacy in smart building management systems.

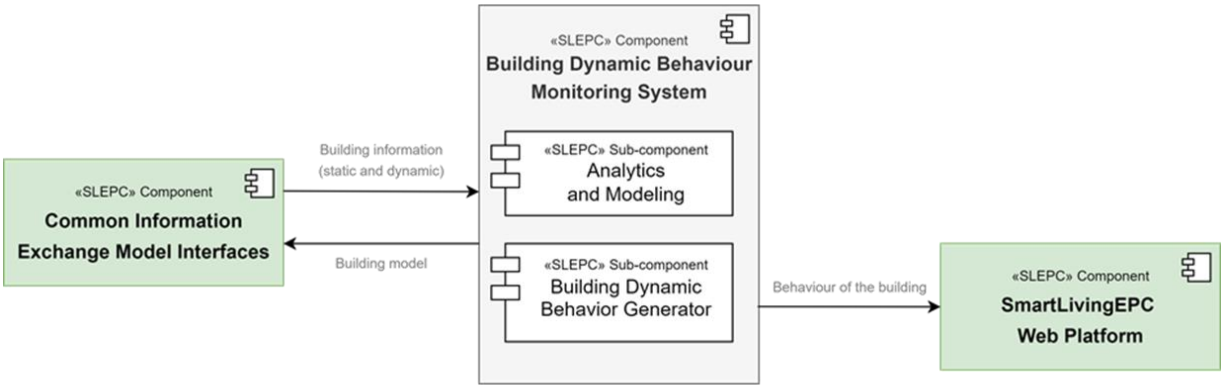

Figure 16: Bulling Dynamic Behaviour Monitoring System functional diagram

6.1 Analytics and Modelling

The Analytics and Modelling phase is a fundamental segment of Building Dynamic Behaviour Monitoring Systems methodology developed by CERTH. It leverages data to illuminate the complex workings of building systems. The subsequent steps aim to craft models that predict and enhance building operations, driving towards energy efficiency and operational excellence.

Specifically:

- **Data Retrieval and Pre-processing**: In this stage, data was acquired from the nZEB Smart House repository. The data covered diverse building attributes including energy consumption, presence, temperature, humidity, luminance, CO_2 levels, HVAC status, and fan speed. The collected dataset possessed a temporal resolution, ranging from high-frequency 5-minute intervals to hourly data points.

- **Correlation and Insight Generation**: After gathering the data, the subsequent step involved establishing correlations between different attributes. This entailed detailed analysis to unveil patterns and dependencies between variables like temperature and energy consumption or presence and lighting status. The objective was to create a model capable of accurately extracting presence in the building and comprehending the dynamic relationships among various environmental parameters.
- **Dataset Resolution Optimization**: Datasets were created with varying time resolutions to optimize the subsequent machine learning models' performance. These ranged from high resolution (5-minute intervals) to lower resolution (60-minute intervals), each accompanied by corresponding sets of data for the training and testing phases.
- **Classifier Model Development:** Different machine learning classifiers were explored to ascertain the best fit for the data and system objectives. These encompassed Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbours (KNN), Decision Tree Classifier (DTC), XGB Classifier, Linear Discriminant Analysis (LDA), and Logistic Regression. These models were utilized to classify and predict the building's dynamic behavior.
- **Data Processing Tools and Code Optimization**: Tools such as MinMaxScaler and Kfold were employed to standardize the dataset and optimize the splitting of data for model validation, respectively. Cross-validation scores and pipelines were used to optimize the machine learning code, ensuring that the models are both accurate and efficient.
- **Accuracy Enhancement Steps**: To enhance the accuracy of the models, the approach included steps such as data scaling with MinMaxScaler to normalize the feature scale, data splitting optimization using Kfold, resolution adjustments, and feature extraction techniques like extracting the humidity ratio from the relative humidity and temperature.

Each step in the Analytics and Modelling phase was carefully designed to build a comprehensive understanding of the building's behaviour through data-driven insights. The methodology emphasizes the use of advanced data analytics, machine learning models, and optimization techniques to achieve high accuracy in behaviour prediction and energy usage insights.

6.2 Building dynamic behaviour generator

The Building Dynamic Behaviour Generator is an integral part of our comprehensive approach to monitoring and optimizing building operations. It focuses on profiling and predicting occupancy patterns, creating a nuanced understanding of building usage. By turning data into visual insights and validating models against real-world occupancy, this component enhances the precision of our building management systems.

Specifically, Building Dynamic Behaviour includes:

- **Dynamic Behaviour Profiling**: This involves understanding and documenting the patterns of occupancy and energy usage within the building over time. By profiling these patterns, it is possible to generate insights into how a building is used and where efficiencies or improvements can be made.
- **Occupancy Estimation Modelling**: Occupancy plays a crucial role in building dynamics. Here, models were designed to estimate occupancy levels using data like CO_2 concentration, motion detection, and energy consumption. Advanced algorithms such as K-Means clustering and Decision Tree Classifiers help to categorize occupancy levels and predict future patterns.
- **Data Visualization:** The data should be presented in intuitive formats such as scatter plots and decision trees to visualize the clustering of occupancy behaviours and the decision rules based on building data features. This helps in understanding complex relationships and contributing factors to building dynamics.
- **Survey and Ground Truth Validation**: In order to validate the occupancy estimation models, a survey was conducted to establish a ground truth dataset. This dataset provides an accurate representation of actual occupancy, which is essential for calibrating the models and ensuring their predictive accuracy.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

The Dynamic Behaviour Generator is an integral component that provides actionable insights through data analysis and modelling, enabling stakeholders to make informed decisions for building management and optimization.

7 Building Dynamic Behaviour Monitoring System: Algorithms

The following sections elucidate the methodologies applied in the preprocessing and analysis of data pivotal to enhancing Building Dynamic Behaviour Monitoring Systems. Initially, Section 6.1 discusses the preparatory steps—scaling, optimizing data splitting, and resolution enhancement—critical for subsequent analytical accuracy. Feature extraction, particularly the calculation of humidity ratio, is also addressed as a key step in enriching the dataset. Subsequently, Section 6.2 delves into the analytics and modelling techniques utilized for occupancy estimation and profiling within the system. The presentation covers also the utilization of diverse classifiers, their performance across datasets of different time resolutions, and the application of clustering algorithms to identify occupancy patterns. The resulting models' performance are visualized through figures and tables, offering insights into the predictive capabilities of the system.

7.1 Data Pre-processing

In Section 6.1, the initial stages of data pre-processing were introduced—a necessary step that laid the foundation for accurate and effective modelling. Techniques such as data scaling using the MinMaxScaler, optimization of data splitting through Kfold cross-validation, and resolution enhancement via upsampling were thoroughly explored. These processes are crucial for ensuring data quality and reliability before undertaking any substantial analysis. Specifically:

Data Scaling (MinMaxScaler) [48]

Data scaling is a crucial pre-processing step in machine learning that involves transforming the range of feature values to ensure that no single feature dominates the model's learning process. The MinMaxScaler is a tool often used for this purpose; it rescales the data set such that all feature values are within a given range, typically 0 to 1. This process preserves the relationships among the original data values while making the model less sensitive to the scale of features.

Data Splitting Optimization (Kfold) [49]

The Kfold method is a cross-validation technique used to assess the predictive performance of a model and ensure that it is not overfitting to the training data. This method involves dividing the dataset into 'k' subsets. In each round of cross-validation, a different subset was held out as the test set, while the remaining subsets were used for training. This cycle was repeated until each subset had been used as the test set. This technique is beneficial because it uses all data points and provides a robust estimate of the model's performance on an independent dataset.

Resolutions (Upsample) [50]

In time-series data analysis, resolution refers to the granularity of data points captured over time. To align datasets with different temporal resolutions, upsampling was used. Upsampling in Python increases the frequency of the data points, interpolating additional points between the existing ones to achieve a higher resolution. This process is essential when merging multiple time-series datasets that have been recorded at different intervals to ensure a consistent time frame for analysis.

Feature Extraction (Humidity Ratio from Relative Humidity and Temperature)

Feature extraction is a process of deriving new features from the existing data which are significant for making predictions. In the context of a smart house monitoring system, extracting the humidity ratio from relative humidity and temperature is an example of creating a new feature that could be more predictive of occupancy

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

or energy consumption than the raw measurements alone. This new feature combines two different measurements to provide a single, more informative value, potentially improving the model's accuracy.

7.2 Implementation

7.2.1 Analytics and Modelling: Model training (Occupancy Estimation)

In this section, a series of charts are presented in Figure 18, Figure 19, Figure 20 and Figure 21, each directly showing the results of subcomponent «Analytics and Modelling» through various classifiers across four datasets differentiated by their time resolutions. These charts effectively demonstrate how the modelling has been applied and the outcomes it has yielded, with specific focus on datasets with 5-minute, 15-minute, 30-minute, and 60-minute granularity. This detailed representation allows for a comprehensive comparative analysis, highlighting the impact of data resolution on the performance of the classification algorithms. Particular attention was paid to metrics such as Accuracy, F1 Score, and Precision, which collectively provide a multi-faceted understanding of each classifier's performance. Specifically:

Accuracy: This metric calculates the proportion of true results (both true positives and true negatives) among the total number of cases examined. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

F1 Score: The F1 score is the harmonic mean of precision and recall. It takes both false positives and false negatives into account and is defined as:

$$F1\ score = \frac{TP}{TP + \frac{1}{2(FP + FN)}} \tag{2}$$

Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It is a measure of a classifier's exactness. Precision is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Where (See Confusion matrix[26] in Figure 17):

- *TP* (True Positives) are the correctly predicted positive values.
- *TN* (True Negatives) are the correctly predicted negative values.
- *FP* (False Positives) are the values which were actually negative but predicted as positive.
- *FN* (False Negatives) are the values which were actually positive but predicted as negative.

[26] A confusion matrix is a table often used to describe the performance of a classification model on a set of data for which the true values are known. Each row of the matrix represents the instances in an actual class, while each column represents the instances in a predicted class.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

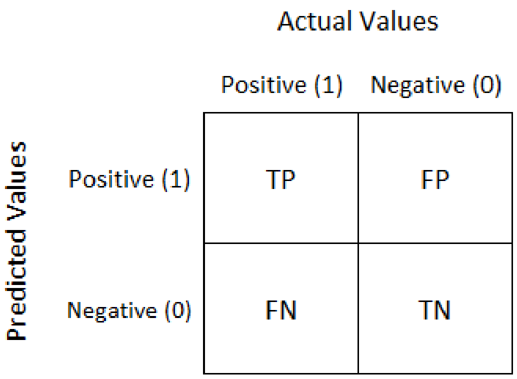

Figure 17: Confusion Matrix

These metrics, ranging from 0 to 1, serve as a benchmark for assessing the classifiers, with higher scores indicating superior performance.The intricate balance between Precision and the F1 Score is especially pivotal in scenarios with uneven class distribution or where the cost of False Positives is high. In contrast, accuracy emerges as a key metric in situations where class distributions are balanced and the consequences of misclassification are symmetric. Through this comprehensive analysis, we aim to uncover the optimal classifier and data resolution pairing, ultimately guiding effective and efficient predictive modelling in varying scenarios.

As already mentioned Figure 18, Figure 19, Figure 20 and Figure 21 demonstrate the outcomes of the "Analytics and Modelling" subcomponent of the Building Dynamic Behaviour System using different classifiers on four datasets distinguished by their time resolutions. Each classifier has three bars corresponding to the three aforementioned metrics (i.e, accuracy, F1 score, precision). The height of the bar indicates the performance value on a scale from 0 to 1, with 1 being the best possible score.

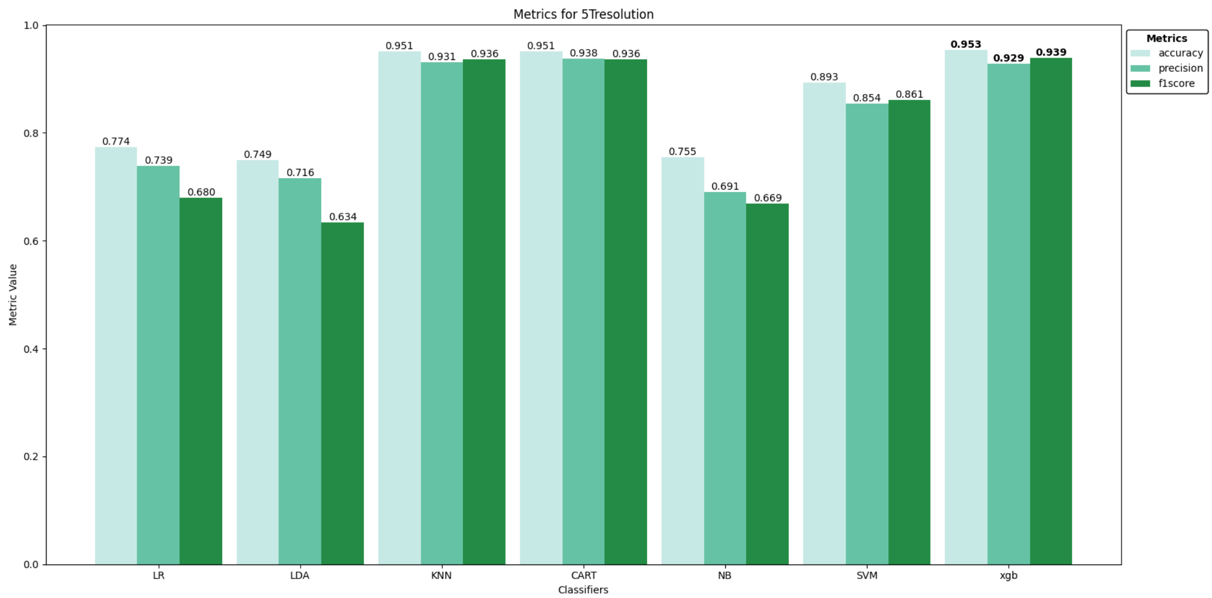

Figure 18: Models for 5 min resolution dataset

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

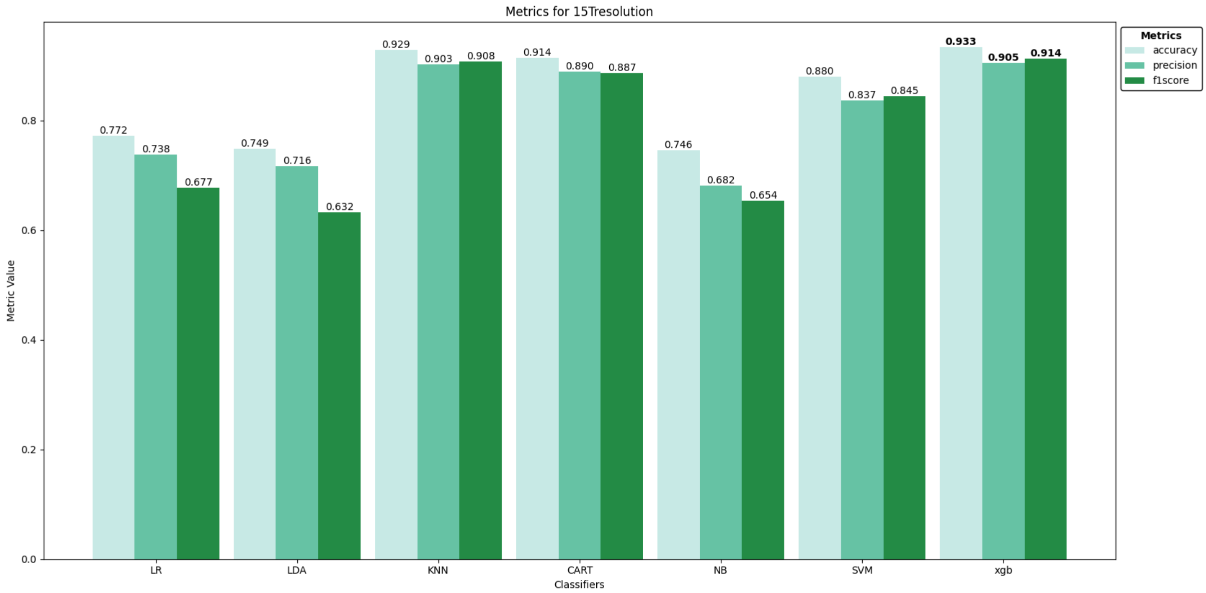

Figure 19: Models for 15 min resolution dataset

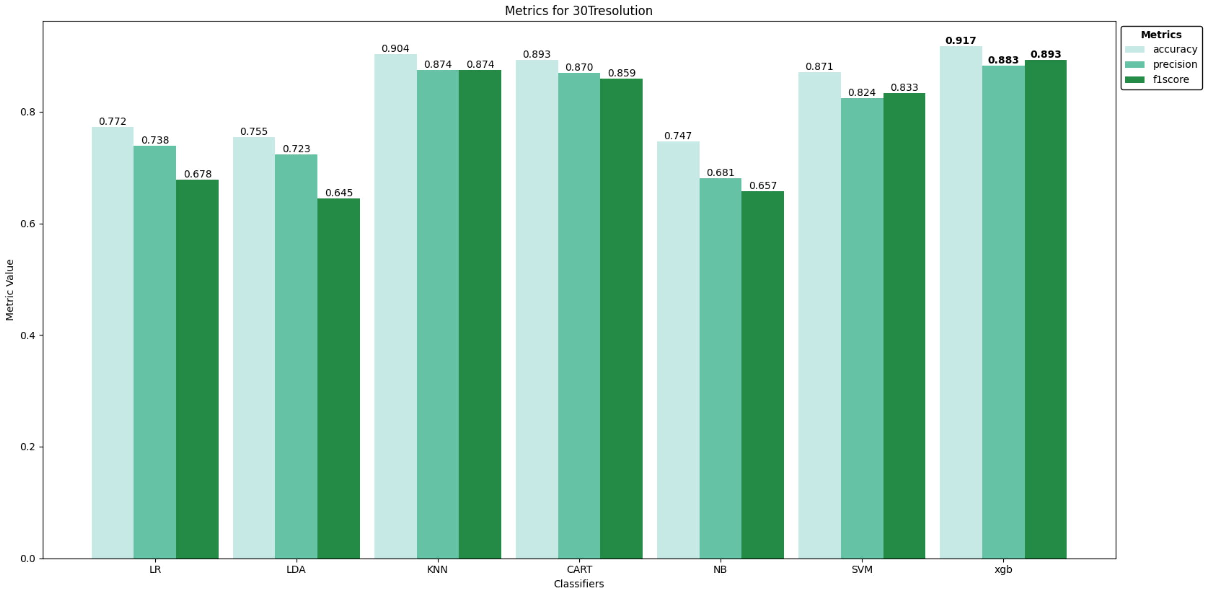

Figure 20: Models for 30 min resolution dataset

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

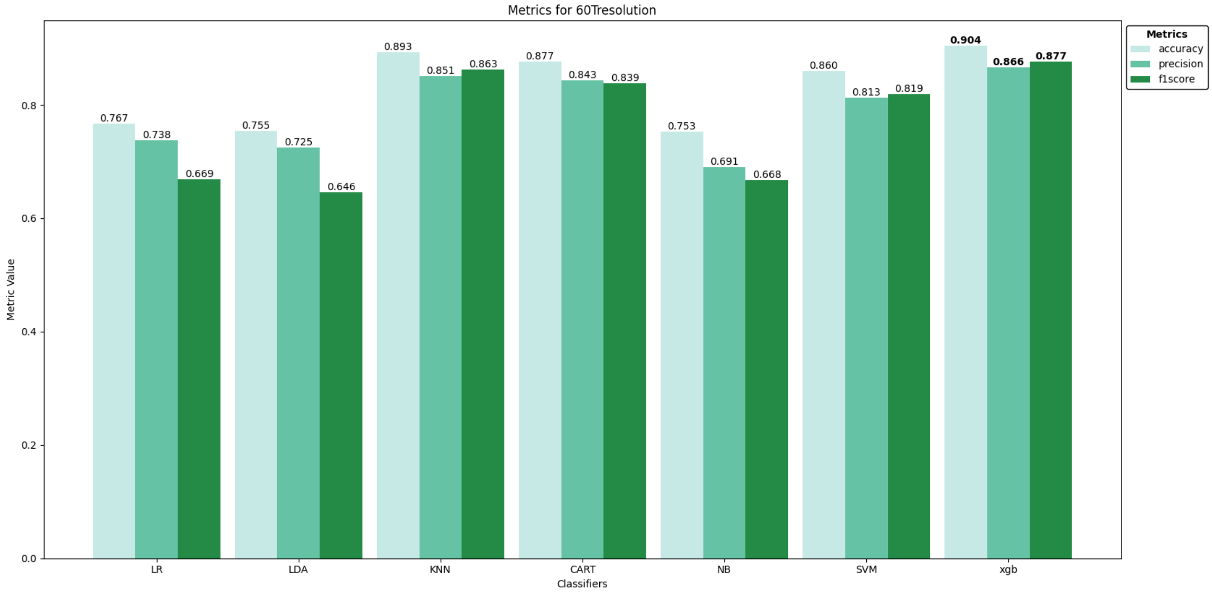

Figure 21: Models for 60 min resolution dataset

Based on the observed data in Figure 18, Figure 19, Figure 20 and Figure 21, it can be stated that the **xgb classifier** consistently outperformed others across all evaluated metrics, closely followed by the CART classifier as a strong contender. Overall, a discernible pattern emerges where the CART and xgb classifiers demonstrate superior performance relative to their counterparts.

Table 4: Confusion Matrix for 5 minutes resolution dataset

Classifier	True Negative	False Positive	False Negative	True Positive
Xgb	**15882**	**748**	**509**	**9742**
CART	15598	632	676	9375
kNN	15577	713	608	9643
SVM	15077	1513	1366	8885

7.2.2 Building dynamic behaviour generator: Model training (Occupancy Profiling)

In this section, the relationship between energy consumption, occupancy, and HVAC status was firstly analysed with respect to the indoor temperature, utilizing the Building Dynamic Behaviour Generator sub component as a pivotal tool in our analysis. The next step was the extraction of the occupancy levels of a building. The process began with importing and pre-processing relevant data attributes such as CO2 levels, motion detection, and energy consumption. Afterwards, the K-Means clustering algorithm was applied to the normalized data to identify **patterns and group data** points into four distinct clusters as we can see in Figure 24. This assumption of four clusters was based on the preliminary analysis suggesting that such a number would optimally represent different occupancy levels. In deeper research for the occupancy levelling, a Decision Tree Classifier was created to classify data points into one of the four clusters. This classifier was deliberately limited to a maximum of four-leaf nodes to correspond with the number of clusters identified by the K-Means algorithm. It was trained on the normalized dataset to predict the cluster labels accurately.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

Figure 22 illustrates that energy consumption in Demo Site #1: nZEB Smart House, aligns with occupancy levels throughout the day but shows a premature decrease before the occupancy drops. This earlier reduction in energy consumption can be explained by Figure 23, which displays the HVAC status. It is observed that the HVAC system's activity starts to decline as the indoor temperature rises, indicating a reduced need for heating. This relationship suggests that the HVAC system is responsive to the ambient temperature, leading to energy savings by decreasing its activity when the demand for warmth lessens. This analysis underscores the HVAC system's role in energy efficiency, particularly in relation to occupancy patterns and temperature changes within the building.

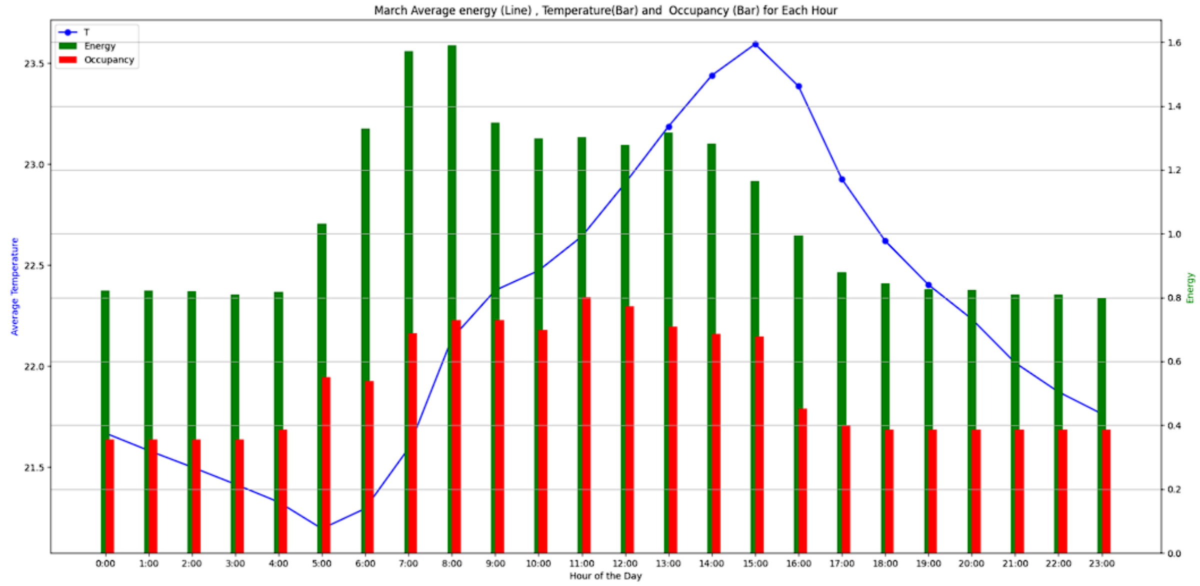

Figure 22: Average energy (Line), Temperature (Bar) and Occupancy (Bar) for each hour

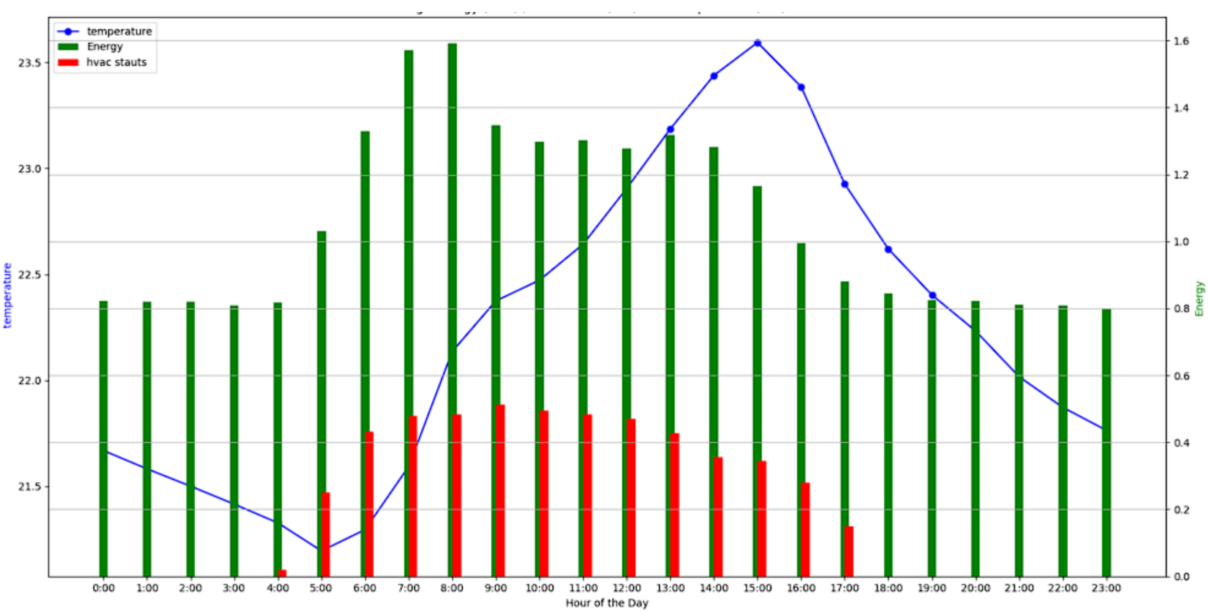

Figure 23: Average energy (Line), HVAC status (Bar) and temperature (Bar) for each hour

In Figure 24, the normalized data underwent application of the K-Means clustering algorithm, revealing distinct patterns and categorizing data points into four separate clusters.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

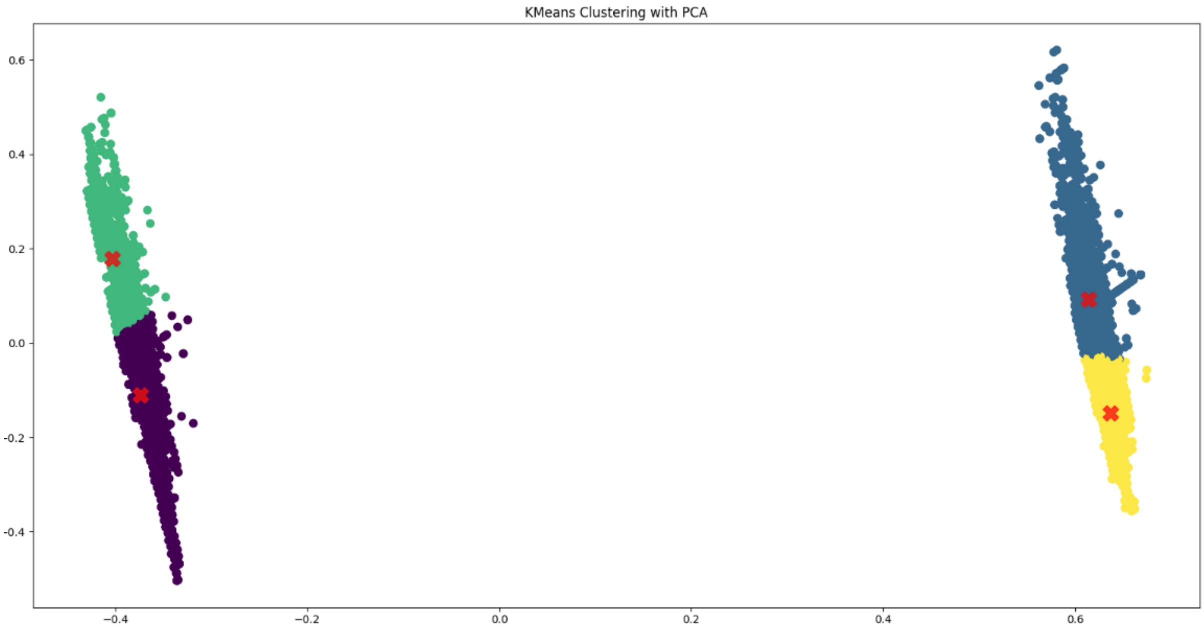

Figure 24: Kmeans clustering

The decision tree (see Figure 25) utilized in the occupancy estimation model prioritizes features based on their predictive power, with splits occurring first on the 'motion detection' feature and subsequently on 'CO2' levels. This indicates the model's reliance on these features as primary indicators of occupancy within the space. The 'gini' values present in the tree's nodes reflect the purity of each cluster, suggesting how homogenous the data points within each node are. A lower 'gini' value would imply a more homogenous cluster. The 'samples' value in each node denotes the count of data points that fall within that specific node, providing insight into the cluster's size and the distribution of data across different clusters.

The leaf nodes of the decision tree signify the final cluster assignments, with the 'value' array contained in these nodes depicting how data points from the original dataset are distributed across these clusters. These clusters are representative of varying levels of occupancy, and by analysing the 'value' arrays, one can infer the predominant level of occupancy associated with each cluster. This mechanism facilitates the understanding of occupancy patterns, serving as a foundational element for smart systems designed to optimize energy usage based on occupancy.

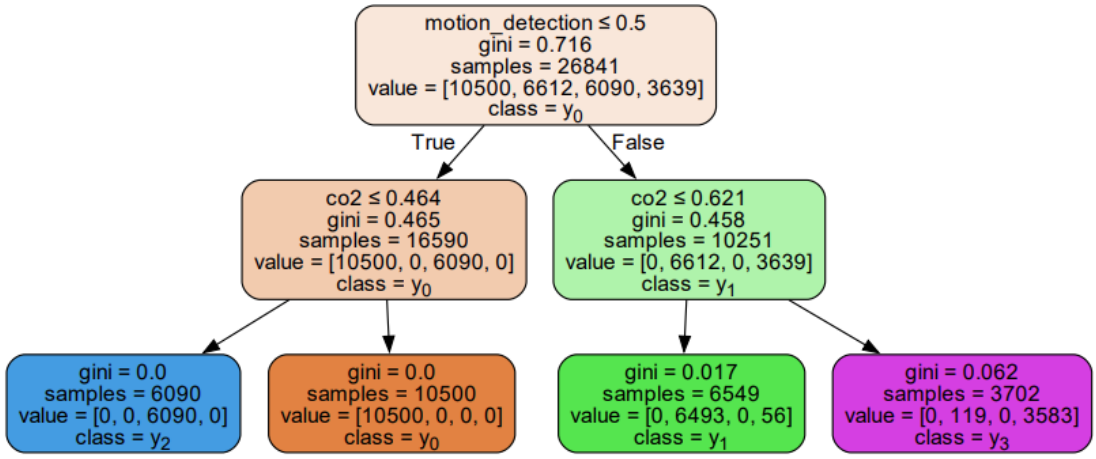

Figure 25: Decision Tree

A questionnaire were conducted to the occupants of Demo Site #1: nZEB Smart House, to understand the patterns of building occupancy over a week. Its extension to a longer period is essential to gain a more thorough and detailed insight into how occupancy patterns vary over time. For this extended duration, stochastic algorithms will be used to generate occupancy data, providing a broader perspective. Following this phase, the integration of machine learning algorithms is planned to be used for refining estimates of the number of individuals in the building at different times, significantly improving the precision and effectiveness of the occupancy profiling.

8 Conclusions and Next steps

The comprehensive analysis presented in this report outlines the vital components of the Common Information Exchange Model (CIEM) and of the Building Dynamic Behaviour Monitoring System. In the present deliverable, the current status of the development phase of these two SmartLivingEPC components is documented.

In alignment with the activities of T4.1, a first version of the CIEM, which serves as a data backbone, enabling seamless exchange of information between the SmartLivingEPC components and the various data sources, is presented. Facilitating the aggregation, normalization, and dissemination of data from diverse sources, CIEM ensures that the modelling and analytics components operate on accurate and comprehensive datasets. It underscores the value of standardized data exchange in enhancing the predictive accuracy and operational efficiency of the proposed system.

As we navigate from the design phase of the SmartLivingEPC project towards the development phase, the conceptualization of the CIEM, remains a pivotal focus. The literature review conducted indicates that the design of CIEM, leveraging semantic web technologies, constitutes a necessary aspect of the CIEM development towards the realization of a unified BIM-IoT data management system, that acts as the middleware for real-time data exchange. Note that, while the current focus has primarily been on constructing and optimizing the data model for individual buildings, it represents a foundational milestone in the project’s larger endeavour encompassing building complex-scale data requirements (assessed in the context of WP2 and WP3). The first steps towards the data model generation have been established and the lifecycle and interaction between ontologies and data models have been defined.

Our immediate trajectory (within M19) involves refining CIEM’s data model to ensure its robustness at the building level before extending its capabilities to encompass the broader complexities of our project's data requirements, i.e., building complex level. By strategically incorporating semantic web technologies, it is anticipated that a more expansive and versatile CIEM implementation, seamlessly integrating diverse data sets across various levels of the project (building and complex) shall be ready by M21. This evolution will amplify SmartLivingEPC’s efficiency, laying a robust foundation for comprehensive fused BIM-IoT data management and analysis, thereby steering us closer to the project’s overarching objectives. Eventually, following the integration of the CIEM with the pilots and based on its performance to effectively and seamlessly retrieve, manage, store and share data with the various SmartLivingEPC components, CIEM will be further extended and refined to finally deliver its second and final version in M22.

As regards T4.2 activities, data preprocessing and optimization, employed MinMaxScaler and Kfold methods to refine and prepare the data for training the model. The implementation of upsampling techniques ensured consistency in time-series data analysis across different resolutions, augmenting model accuracy. Additionally, ingenious feature extraction methods, such as deriving the humidity ratio from temperature and relative humidity, combined with sophisticated analytical algorithms, have enriched predictors for occupancy estimation. Also, advanced algorithms, notably K-Means and Decision Trees, have been utilized for occupancy profiling, rigorously evaluated across varied datasets to ascertain optimal model performance. Visual representations, including scatter plots and decision tree visualizations, have provided insights into the correlation between occupancy, energy consumption, and HVAC status, elucidating the building's operational dynamics. Moreover, the analysis has revealed a significant correlation between indoor temperature fluctuations and HVAC activity, suggesting potential energy-saving opportunities. The utilization of a diverse ensemble of trained models, particularly the robust xgb and CART classifiers, has proven invaluable in comprehending occupancy patterns essential for optimizing energy usage. These completed actions collectively contribute to a more nuanced understanding of building behaviour and pave the way for optimizing energy efficiency.

Future steps in the area of occupancy profiling will be approached through the development of a Building Dynamic Behavior Monitoring System. Initially, research that has been conducted in the building dynamic behaviour generator, aimed at understanding occupancy patterns over a week, is to be expanded over a longer period. This extension is necessary for a more detailed and comprehensive understanding of the variations in occupancy over time. Stochastic algorithms are to be utilized for generating occupancy data for this extended duration, enabling insights that extend beyond the initially observed week by the end of M20. Following this, the

The SmartLivingEPC project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101069639.

integration of machine learning algorithms is planned. These algorithms will be used to refine the estimates of the number of people within the building at different times. This phase is critical for improving the precision and efficacy of the occupancy profiling. Also, the Analytics and Modelling will be trained again with data from one year. Updates for each of the 2 sub-components will be completed by the end of M21. Finally, the successful integration of the Building Dynamic Behaviour Monitoring System will be expanded and refined in order to be delivered in the final version of this deliverable by M22.

References

[1] T. Roshana, H. Mohd and N. Abdul Hadi, "Building Information Modeling (BIM): A new paradigm for quality of life within Architectural, Engineering and Construction (AEC) industry," *Procedia - Social and Behavioral Sciences,* vol. 101, pp. 23-32, 2013.

[2] J. Shaohu, J. Liping, H. Yunwei, W. Zheng and W. Na, "OpenBIM: An Enabling Solution for Information Interoperability," *Applied Sciences,* vol. 9, no. 5358, 2019.

[3] I. Alothaimeen, D. Arditi and O. H. Türkakın, “Multi-objective optimization for LEED - new construction using BIM and genetic algorithms,” *Automation in Construction,* p. 104807, 2023.

[4] S. Fenz, G. Giannakis, J. Bergmayr and S. Iousef, “RenoDSS-A BIM-based building renovation decision support system,” *Energy and Buildings,* p. 112999, 2023.

[5] S. Seungwoo, L. Ghang, J. Jeaeun, K. Jungdae and J. Kahyun, "Automated generation of a model view definition from an information delivery manual using idmXSD and buildingSMART data dictionary," *Advanced Engineering Informatics,* p. 101731, 2022.

[6] A. Andriamamonjy, D. Saelens and R. Klein, “An automated IFC-based workflow for building energy performance simulation with Modelica,” *Automation in Construction,* pp. 166-181, 2018.

[7] D. Kontaxis, G. Tsoulos, G. Athanasiou and G. Giannakis, "Wireless Sensor Networks for Building Information Modeling," *Telecom,* vol. 3, pp. 118-134, 2022.

[8] X. Huang, Y. Liu, L. Huang, E. Onstein and C. Merschbrock, "BIM and IoT data fusion: The data process model perspective," *Automation in Construction,* vol. 149, p. 104792, 2023.

[9] H. M. Baydaa, S. Hasimi, S. Nurhizam, H. Afifuddin, A. B. Nur Azaliah, Y. Farashazillah, A. Ihsan and A. M. Shaymaa, "Building Information Modeling and Internet of Things Integration in the ConstructionI ndustry: A Scoping Study," *Advances in Civil Engineering,* pp. 1-20, 2022.

[10] Abubaker, Basheer Abdalwhab Altohami; Nuzul, Azam Haron; Aidi, Hizami Ales; Teik, Hua Law, "Investigating Approaches of Integrating BIM, IoT, and Facility Management for Renovating Existing Buildings: A Review," *Sustainability,* p. 3930, 2021.

[11] S. Tang, D. R. Sheldena, C. M. Eastmana, P. Pishdad-Bozorgib and X. Gao, "A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends," *Automation in Construction,* vol. 101, pp. 127-139, 2019.

[12] A. Donkers, D. Yang, B. de Vries and N. Baken, "Semantic Web Technologies for Indoor Environmental Quality: A Review and Ontology Design," *Buildings,* vol. 12, p. 1522, 2022.

[13] M. Marzouk and A. Abdelaty, "Monitoring thermal comfort in subways using building information modeling," *Energy and Buildings,* vol. 84, pp. 252-257, 2014.

[14] Y. Zhang and L. Bai, "Rapid structural condition assessment using radio frequency identification (RFID) based wireless strain sensor," *Automation in Construction,* vol. 54, pp. 1-11, 2015.

[15] W. Solihin, C. Eastman, Y.-C. Lee and D.-H. Yang, "A simplified relational database schema for transformation of BIM data into a query-efficient and spatially enabled database," *Automation in Construction,* vol. 84, pp. 367-383, 2017.

[16] T.-W. Kang and H.-S. Choi, "BIM perspective definition metadata for interworking facility management data," *Advanced Engineering Informatics,* vol. 29, pp. 958-970, 2015.

[17] G. Basak et al., "Interoperability in building information modeling for AECO/FM industry," *Automation in Construction,* vol. 113, p. 103122, 2020.

[18] P. Pauwels and W. Terkaj, "EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology," *Automation in Construction,* vol. 63, pp. 100-133, 2016.

[19] G. Fierro and P. Pauwels, "Survey of metadata schemas for datadriven smart buildings (Annex 81)," CISRO, 2022.

[20] B. Zhong, C. Gana, H. Luo and X. Xing, "Ontology-based framework for building environmental monitoring and compliance checking under BIM environment," *Building and Environment,* vol. 141, pp. 127-142, 2018.

[21] S. Sobhkhiz, H. Taghaddos, M. Rezvani and A. M. Ramezanianpour, "Utilization of semantic web technologies to improve BIM-LCA applications," *Automation in Construction,* vol. 130, p. 103842, 2021.

[22] R. García-Castro, M. Lefrançois, M. Poveda-Villalón and L. Daniele, "The ETSI SAREF Ontology for Smart Applications: A Long Path of Development and Evolution," John Wiley & Sons, Inc, 2023.

[23] K. Amasyali et al., "A review of data-driven building energy consumption prediction studies," *Renewable and Sustainable Energy Reviews,* pp. 1192-1205, 2018.

[24] X. Xu et al., "Incorporating machine learning with building network analysis to predict multi-building energy use," *Energy & Buildings,* 2019.

[25] M. Khalil et al., "Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review," *Engineering Applications of Artificial Intelligence,* 2022.

[26] X. Luo et al., "Comparative study of machine learning-based multi-objective prediction framework for multiple building energy loads," *Sustainable Cities and Society,* 2020.

[27] S. Seyedzadeh et al., "Machine learning modelling for predicting non-domestic buildings energy building energy consumption—A systematic review," *Applied Energy 279 (2020) 115908,* 2020.

[28] H. Hosamo Hosamo et al., "Multiobjective optimization of building energy consumption and thermal comfort based on integrated BIM framework with machine," *Energy & Buildings,* 2022.

[29] P. Wenbin Tien et al., "Real-time monitoring of occupancy activities and window opening within buildings using an integrated deep learning-based approach for reducing energy demand," *Applied Energy,* 2022.

[30] W. Li et al., "A novel operation approach for the energy efficiency improvement of the HVAC system in office spaces through real-time big data analytics," *Renewable and Sustainable Energy Reviews,* 2020.

[31] M. Esrafilian-Najafabadi et al., "Occupancy-based HVAC control systems in buildings: A state-of-the-art review," *Building and Environment,* 2021.

[32] Y. Yuan et al., "Occupancy Estimation in Buildings Based on Infrared Array Sensors Detection," *IEEE Sensors Journal,* vol. 20, no. 2, pp. 1043-1053, 2020.

[33] Z. Chen et al., "Building occupancy estimation and detection: A review”, in Energy and Buildings," *Energy and Buildings,* vol. 121, pp. 277-290, 2017.

[34] Zhaoxuan Li et al., "A new modeling approach for short-term prediction of occupancy in residential buildings," *Building and Environment,* vol. 121, 2017.

[35] Z. Li, "A new modeling approach for short-term prediction of occupancy in residential buildings," *Building and Environment,* vol. 121, pp. 277-290, 2017.

The SmartLivingEPC project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101069639.

[36] Y. Yuan et al., "Occupancy Estimation in Buildings Based on Infrared Array Sensors Detection," *IEEE Sensors Journal,* vol. 20, no. 2, pp. 1043-1053, 2020.

[37] Y. Penga et al., "Using machine learning techniques for occupancy-prediction-based cooling," *Applied Energy,* no. 211, pp. 1343-1358, 2018.

[38] Z. Chen et al., "Deep Learning for Building Occupancy Estimation Using Environmental Sensors," *Deep Learning: Algorithms and Applications.Studies in Computational Intelligence,* vol. 865, p. 335–357, 2019.

[39] Z. Chen et al., "Building Occupancy Estimation with Environmental Sensors via CDBLSTM," *IEEE Transactions on Industrial Electronics,* vol. 64, no. 12, pp. 9549-9559, 2017.

[40] T. Chen et al., "Xgboost: A scalable tree boosting system.," in *22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016.

[41] Z. Khorasani zadeh et al., "Optimizing occupant-centric building controls given stochastic occupant behaviour," in *Journal of Physics: Conference Series*, 2021.

[42] Ignacio Benítez et al., " Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers " *Electrical Power and Energy Systems,* vol. 55, pp. 437-448, 2014.

[43] K. Prabhakaran et al., "Explainable K-Means Clustering for Occupancy Estimation," in *The 12th International Conference on Sustainable Energy Information Technology (SEIT)*, Niagara Falls, Canada, 2022.

[44] K. Vassiljeva et al., "Data-Driven Building Occupancy Profile Identification and Application for Ventilation Operating Schedule Optimization," 16 Jun 2023.

[45] C. Eastman et al., "Automatic rule-based checking of building designs," *Automation in Construction,* vol. 18, p. 1011–1033, 2009.

[46] X. Zhao et al., "Compliance Checking on Topological Spatial Relationships of Building Elements Based on Building Information Models and Ontology," *Sustainability,* vol. 15, no. 14, p. 10901, 2023.

[47] " UC Irvine Machine Learning Repository," [Online]. Available: https://archive.ics.uci.edu/. [Accessed 19 5 2023].

[48] "scikit-learn," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html.

[49] "scikit-learn," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html.

[50] "pandas documentation," [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.resample.html.

[51] J. Smith, "inteGRIDy sample reference," *EU,* pp. 1-2, 2016.

[52] "ETSI TS 103 410-3 V1.1.2: SmartM2M; Extension to SAREF; Part 3: Building Domain," 2020.

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.

HE Grant Agreement Number: 101069639
Document ID:
WP<n.>/D/IR

Advanced Energy Performance Assessment towards Smart Living in Building and District Level

https://www.smartlivingepc.eu/en/

https://www.linkedin.com/company/smartlivingepc/

https://twitter.com/SmartLivingEPC

https://www.youtube.com/channel/UC0SKa-20tiSabuwjtYDqRrQ

The SmartLivingEPC project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101069639.