3-A-4-양-2
SCIE

Client-centered detached modular housing: natural language processing-enabled design recommender system

- Journal of Computational Design and Engineering -

2024.11.

과 제 명	인공지능 기반의 건축설계 자동화 기술개발		
주 관 기 관	경북대학교 산학협력단		
총 연 구 기 간	2021. 04 . 01 - 2025. 12 . 31(4년 9개월)		
해당연도(4차년)	2024. 01 . 01 - 2024. 12 . 31(1년)		
구 성 기 술 명	구성기술 3	설계 품질검토 자동화를 위한 지능형 설계서비스 보급·활용 기술개발	
세 부 과 제 명	3-A	지능형 설계적법성 평가 및 건축행정 서비스 지원 기술개발	
공 동 연 구 기 관	경희대학교 산학협력단, ㈜코스펙이노랩		
연 구 기 관	경희대학교 산학협력단	연구책임자	김인한

Journal of Computational Design and Engineering, 2024, **11**, 137–157

DOI: 10.1093/jcde/qwae041

Advance access publication date: 29 April 2024

Research Article

Client-centered detached modular housing: natural language processing-enabled design recommender system

Inhan Kim[1], Jihye Shin[2], Syed Haseeb Shah[1] and Saddiq Ur Rehman[1,*]

[1]Department of Architecture, Kyung Hee University, Yongin-si 17104, Korea

[2]Department of Infrastructure Engineering, The University of Melbourne, Parkville VIC 3052, Australia

*Correspondence: saddiqurrehman@khu.ac.kr

Abstract

Designing modular housing is a complex task that necessitates a thorough understanding of the diverse needs of clients in terms of both design aesthetics and floor plan layout. Furthermore, adhering to design for manufacture and assembly (DfMA) principles adds to the complexity, as these are essential modular home requirements. Traditional construction methods frequently fail to meet the specific needs of both clients and DfMA, potentially resulting in suboptimal design solutions. Incorporating client requirements during the design phase necessitates the use of an effective system and framework to reduce changes in subsequent project stages. Existing literature lacks a suitable approach, particularly in the context of modular housing. To address this gap, this paper introduces an artificial intelligence–building information modeling recommender system (RS) for detached modular housing design. The system processes client requirements entered as text utilizing the Word2vec algorithm with the GloVe dataset, refined through transfer learning using surveyed client data of housing needs. The system recommends three distinct modular building design alternatives sourced from a building information modeling models database using cosine and Euclidean similarity functions. A sensitivity analysis ensures that client needs are considered fairly, increasing the robustness of the RS. By incorporating natural language processing, this system transforms the construction industry by making initial designs more client-centric compared with traditional methods. Furthermore, it promotes improved collaboration among clients, design, and construction teams, reducing modifications to design in later stages of construction.

Keywords: detached modular housing, Word2vec algorithm, building information modeling, design recommendation system, client needs, similarity functions

1. Introduction

Detached modular housing has received significant attention over the years due to the advantages it offers in comparison with traditional construction methods. The process of detached modular housing consists of a combination of controlled factory settings and modular construction which allows for unparalleled flexibility and efficiency in building design and manufacturing. The factory-controlled environment ensures the quality and consistency of each module, leading to a superior product. By leveraging these benefits, organizations can achieve their building objectives in a cost-effective, dependable, and sustainable manner. With this innovative approach, buildings can be assembled with precision and speed, resulting in streamlined construction timelines, and minimized on-site disruption (Bertram *et al.*, 2019; Lawson *et al.*, 2012; Pan & Sidwell, 2011; Velamati, 2012). Using modular construction, the speed of a project can be increased by 20% to 50% and with proper environment and trade-offs, 20% cutoff in project budget can be achieved (Bertram *et al.*, 2019). Modular construction is suitable for a wide range of building types, including residential, commercial, industrial, and institutional buildings (Modular Building Institute, 2021; National Institute of Building Science, 2018). Despite its numerous benefits, modular construction is still uncommon in the construction industry due to various barriers such as lack of awareness, stakeholder acceptance, cost-effectiveness concerns, design limitations, logistical issues, and the absence of integration frameworks and specific guidelines (Abd Razak *et al.*, 2022). Highlighting the importance of comprehensive strategies (Wuni & Shin, 2020), frameworks such as design for manufacture and assembly (DfMA, Rankohi *et al.*, 2023) and specific elements such as design (Baro *et al.*, 2022; Gbadamosi *et al.*, 2019), cost-efficiency (Almashaqbeh & El-Rayes, 2021), standards (Rehman *et al.*, 2022), and logistics (Lee *et al.*, 2022) are critical for encouraging the shift from traditional construction methods to modular construction.

Modular housing construction offers high levels of customization, which underscores the importance of early client engagement due to the limited flexibility in the post-production modification. Similar to traditional projects, successful modular construction projects rely heavily on client collaboration. A deep understanding of client needs not only elucidates the requirements for the final product but also optimizes the design for efficiency and customization, streamlining the construction process and enhancing both its efficiency and cost-effectiveness (Designing Buildings, 2022). This customization encompasses changes to

Received: January 21, 2024. **Revised:** April 15, 2024. **Accepted:** April 15, 2024

various elements, including floor plans, layout, structure, facade, and finishes, to bring the client's vision to life, all while considering the constraints of post-production adaptability (Hentschke *et al.*, 2014). The controlled factory environment ensures precise module fabrication but makes subsequent modifications difficult and expensive (Bertram *et al.*, 2012). Therefore, key decisions regarding modular production, installation, and interface planning must be made during the early design phase to align modules with specific client requirements, emphasizing the need for a dynamic, client-centric approach to modular construction (Abdul Nabi & El-adaway, 2020; Rehman *et al.*, 2023). Achieving this requires specialized design configuration knowledge for modular housing. Architects must have a comprehensive understanding of modular house production and assembly processes to provide diverse solutions that meet unique client-specific requirements while adhering to building codes (Ghannad & Lee, 2022; Svensson & Barfod, 2002; Westbrook & Williamson, 1993).

The current process of modular housing construction presents challenges in conveying client needs to their project products, affected by various factors. These challenges include the absence of experienced collaborative teams, a limited comprehension of stakeholder' roles, and difficulties in effective communication among stakeholders on the client's part (Kamali & Hewage, 2016; Nguyen *et al.*, 2022). Within traditional structured design process, where client requirements are defined in a brief and architects subsequently translate and elaborate on these requirements in technical terms, modular housing design options have been explored by architects relying solely on their experience, often lacking a systematic analysis of client needs. This approach frequently results in limitations in exploring the full spectrum of modular house project aspects, from initial design to the installation, and necessitates costly and time-consuming iterations of design changes (Brandão, 2011). These distinctions underscore the necessity of a dynamic, client-centric approach to modular housing construction, redefining how client needs are met within this innovative building methodology (Abdul Nabi & El-adaway, 2020). However, client collaboration within offsite construction, particularly during the design phase, has received limited attention in both academic and practical contexts (Othman, 2015). Despite the pivotal role of client involvement in the success of modular projects, there is currently no comprehensive approach available to address client needs in modular design. What is needed is a new approach that can efficiently incorporate client feedback and requirements into the design process and offer customized design options based on the client specific needs and preferences. This approach could reduce the time and effort required to design modular housing while simultaneously improving the quality of the final design solution.

Building information modeling (BIM) is a dynamic process for creating and managing information on a construction project throughout its whole lifecycle. It has been conceived as an enabler of successful collaboration among diverse stakeholders throughout the project lifecycle (Sebastian, 2011; Shin *et al.*, 2020). Effective collaboration requires the ability to input, extract, update, or modify information in accordance with the distinct roles of each involved stakeholder. BIM facilitates this collaboration by enabling the development of a shared and integrated digital representation of multidisciplinary information of building, based on open standards for interoperability. It evolved into a virtual information model that seamlessly transfers from the design team to contractors, subcontractors, and, ultimately, the client (Sebastian *et al.*, 2009). Research has demonstrated that using BIM in the design process results in higher customer satisfaction, underscoring its effectiveness in meeting client needs (Dauphin Americas, 2022). A survey conducted among architecture firms has shown that BIM can be leveraged to create user-centered smart working environments, thus fostering greater client involvement in the design phase (Park *et al.*, 2022). Many studies highlight the potential of BIM as an effective collaborative mechanism for modular building design, ensuring client participation in the design decision-making process regarding building elements (Abanda *et al.*, 2017; Bakhshi *et al.*, 2022; Gbadamosi *et al.*, 2019).

This paper aims to respond to the challenge of addressing client needs in residential modular building design by developing an artificial intelligence–building information modeling (AI–BIM) recommender system (RS). An RS is a tool that analyzes user preferences and behaviors to provide personalized recommendations (Amara & Subramanian *et al.*, 2020). The RS proposed in this study analyzes client requirements, and it suggests suitable modular design options within the BIM environment, employing machine learning algorithms. To accomplish this aim, this research first identifies housing preference of clients through surveys involving 500 individuals who have constructed or intend to construct houses, regardless of whether they choose modular or non-modular construction methods. Of these, the most prominent needs have been derived, using text processing method – Latent Dirichlet Allocation (LDA) topic modeling. Furthermore, this research creates BIM models, each incorporating these prominent needs in various combinations to serve as three-dimensional (3D) digital representation of the needs. We established a BIM database to store and manage these BIM models in systematic manners. Additionally, the client needs are translated into vectors, making them machine-readable, through employment of the Word2vec algorithm, a text processing technique. Moreover, relationships between the client needs in vector form and BIM database schema are identified and quantified using Euclidean similarity measured, forming the core of the RS. Utilizing the similarity measurements, we designed an AI–BIM RS capable of suggesting various house designs from the BIM database based on input needs of clients. The robustness and reliability of the recommendation system has been validated using sensitivity analysis. This system has the potential to enhance client-centric approaches by enabling real-time client interactions and delivery of customized recommendations, resulting in a precise alignment of final design with client expectations as well as heightened client satisfaction. Furthermore, it can facilitate the optimization of the modular construction process, guaranteeing the provision of personalized, efficient design solutions.

The article is organized as follows: Section 2 delves into related literature, highlighting the shortcomings in detailing client needs in modular construction. Section 3 describes the research methodology, focusing on data acquisition process, BIM models and database development process, training process of Word2vec model, and methods used for similarity measurement and validation. Section 4 describes the implementation of artificial intelligence (AI) and BIM RS in which the result of the methodology subsections is presented with the system architecture of the RS. And Section 5 discusses the implication of the study and compares it with existing RS and literature. Finally, Section 6 provides a concise summary of the article, highlighting its conclusions, contributions to the AEC sector, and providing insights for future research.

2. Related Studies

2.1. Client-centered design in modular construction and recommendation systems

In modular construction, client-centered design principles are imperative to ensure that the projects align with specific needs, preferences, and constraints of each client. These principles encompass following key elements: (i) identifying core portfolio elements that cater to customer desires, (ii) understanding the unique planning and design intricacies of modular construction, (iii) facilitating communication through shared models and rapid performance evaluations, and (iv) acknowledging the role of human senses in architectural design (Bertram *et al.*, 2019; El Mounla *et al.*, 2023; Spence, 2020; Wilson, 2019). This approach underscores the critical importance of recognizing and incorporating individual client requirements into the construction process, highlighting the necessity for customized design solutions. Considering client needs during the initial design stages can help mitigate design changes, which has direct impact the overall performance of the project, especially concerning cost, schedule, and quality (Yap & Skitmore, 2018).

Traditional design practices frequently fail to comprehensively meet the specific client needs through appropriate design solutions, resulting in suboptimal design changes. To address this, numerous academics and practitioners have endeavored to develop approaches for considering client needs during the design phases, with the objective of minimizing the necessity for subsequent design modifications. Gao *et al.* (2013) emphasized differences in people's housing preferences and the importance of high-level customization in architectural design. They introduced preference indices to evaluate to what extent people are satisfied with different design options, with particular focus on floor plans. Zawidzki and Szklarski (2020) established a framework that optimizes architectural plan layout for functionality, comfort, and authentic perspectives based on user input preferences on layout and site, utilizing rigorous mathematical optimization methods. In practice, RSs have been considered as a viable solution to addressing client needs within project design (Isinkaye *et al.*, 2015). These systems have proven successful, particularly in managing the challenge of information overload, and have become a favored approach for considering and addressing client needs in a variety of fields such as psychology, mathematics, computer science (Roy & Dutta, 2022; Wei *et al.*, 2005). In the construction industry, several recommendation systems have also been developed. For instance, Architectural Designs (2023), available in the USA and Canada, has been developed to provide design suggestions based on criteria, such as the number of bedrooms, bathrooms, floors, and heated floor area. Similarly, Family Home Plans (Stem, n.d.) has been established to offer housing designs catering to six types of client needs, such as living area, bedroom count, bathroom count, floor levels, garage, and foundation types. Additionally, House Plans (2023), available in the USA, Canada, and Europe, offers a range of design options based on five key client specifications: bedroom count, bathroom count, floor levels, garage count, and total housing area.

Existing studies have shown limited attention to aligning modular housing design with client needs, despite its significance when compared with traditional construction methods. They fall short in considering both the modular aspects and client requirements for building facilities and spaces, instead focusing solely on layout configuration while neglecting facility considerations that should align with client needs. Additionally, the existing recommendation systems tend to suggest building design plans based on filtering plans that meet specific quantitative criteria. They do not fully address intricate design needs and preferences, such as specific layout preferences, aesthetic choices, or unique spatial requirements, which are critical in housing construction, especially modular housing because modular housing presents unique challenges and opportunities due to its prefabricated nature, requiring more emphasis on modularity, adaptability, and assembly considerations in the design process. Aesthetic choices remain critical, with the additional layer of ensuring these preferences are feasible within modular construction constraints. Furthermore, these systems have limitations in leveraging advanced analytics techniques, such as natural language processing or machine learning algorithm, which hampers their ability to offer highly personalized and context-aware design recommendations. Additionally, their deficiency in maintaining a comprehensive database, especially one that integrates client-centric data, could curtail the variety and precision of design options they provide, potentially resulting in suboptimal or generic design solutions that may not align perfectly with individual client needs.

2.2. Client requirements harmonized with BIM in modular construction

The integration of BIM and modular construction has enormous potential in ensuring that the projects align with client needs (Hwang & Kim, 2022). However, meticulous attention to detail during the BIM modeling phase is essential, as is a client-centric approach. Balancing the standardized efficiency of modular construction with the need for customized client solutions is still a significant challenge. It necessitates a careful integration of BIM's digital precision with the adaptability and customization required in modular construction, resulting in a harmonious alignment of digital design and physical modular reality in addressing diverse client requirements. Researchers rarely pay attention to client needs in consideration in the BIM for modular building, their focus is mostly on the manufacturing and fabrication phases. Lu and Korman (2010) highlight the improved interoperability of BIM applications, which improves integrated building design and integrated project delivery. According to the authors, BIM simplifies modular construction by enabling seamless digital design-to-fabrication workflows across disciplines. Nonetheless, challenges remain, such as the need for standardized BIM models and interoperability among various BIM software platforms. To encourage the design and construction of mid-rise modular buildings, Xiao and Watson (2019) propose an information framework and supporting infrastructure. According to the authors, BIM can improve the efficiency and quality of modular construction. However, the authors do not emphasize the need for client needs consideration in the mid-high-rise buildings. Gbadamosi *et al.* (2020) integrate BIM, DfMA, and big data to propose Big Data Design Options Repository (BIG-DOR), a solution that aims to connect BIM clients with manufacturers/suppliers by providing critical data such as prefabricated component costs and production lead times. This integration has resulted in the creation of a design alternatives assessment system, which has increased the use of offsite construction methodologies. However, the design option repository was not developed by preliminary surveying various client needs and creating BIM models based on those needs. The various design alternatives may not be suitable to the requirement of client for modular building design. Nonetheless, the design option repository was not established by first exploring various client requirements and then creating BIM models tailored to these needs. As a result, the variety of design options available may not correspond to the specific requirements of clients for modular building design. Gan (2022) proposed a BIM-based graph data model for the automatic generative design of modular buildings, claiming that the model can generate modular building designs based on user-defined parameters. The graph model was built with a focus on modular building design requirements, with client preferences primarily influencing mass plan development based on specific site conditions. However, critical client needs such as floor plans, interior layouts, and other structural considerations, which frequently drive design changes in later phases, were overlooked. As a result, designs generated by the graph model typically include a mass model that corresponds to client preferences, while other details frequently remain unaligned.

The integration of BIM and modular construction has enormous potential for meeting client needs, but it necessitates meticulous attention to detail and a client-centered approach. Current research frequently overlooks client needs during BIM modeling, instead focusing on the manufacturing and fabrication phases. The BIG-DOR system, e.g., aims to improve offsite construction methodologies but lacks deep client-centric customization. Similarly, the BIM-based graph data model focuses primarily on mass plans, ignoring critical interior layouts and structural considerations important to clients. In the BIM modeling phase, a comprehensive and client-oriented approach is required to align digital precision with client needs in modular construction.

2.3. Architectural design customization through generative algorithms and NLP

A machine learning system draws conclusions from patterns in data and changes its behavior based on the inferences (Alzubi *et al.*, 2018). In today's era, machine learning techniques have found widespread and diverse applications. The AEC industry extensively employs machine learning algorithms in various domains (Yu *et al.*, 2022). On of the earliest examples of machine learning in construction industry is the use of neural networks in 1989 for the design of steel beams (Adeli, 2001). Since then, machine learning has been utilized for various purposes such cost prediction (Wilmot & Mei, 2005), construction management (Heravi & Eslamdoost, 2015), and safety and risk assessment (Li *et al.*, 2020). For design customization, Bianconi *et al.* (2019) used advanced techniques such as parametric modeling and generative adversarial networks (GANs). Their approach includes the development of a web-based catalog capable of generating multiple design options for individual houses. This methodology not only improves critical elements such as functionality, aesthetics, and cost-efficiency, but it also allows architects and clients to work in a more seamless and collaborative manner. However, the incorporation of client preferences occurred only after the architect had chosen a specific design alternative for further detailed development which makes the design alternative proposed by generative design algorithms to be not in compliance with client needs. Another study investigates how Markov decision process (MDP) can be used to improve architectural design processes (Karan & Asadi, 2019). To incorporate subjective data into the architectural design process, the study employs a standard window design experiment based on a design-review-feedback strategy. The MDP algorithm is used in the study to model a window design experiment as a sequential decision problem, yielding favorable results in terms of client satisfaction. However, as stated by the author, the designs proposed by the MDP model may not always align with the client's preferences. So, there is a need to apply AI to design challenges with a greater number of clients needing parameters. Merrell *et al.* (2010) propose an automated method for generating building layouts for computer graphics applications by combining high-level requirements with a Bayesian network trained on real-world data. Based on the architectural program, stochastic optimization is used to generate floor plans, which are then used to build detailed 3D buildings with internal structures. The method is demonstrated using a variety of computer-generated buildings, and it was concluded that it is capable of producing effective applications for collaborative building layout creation. However, it does not take into account the wide range of site-specific and client-specific factors that architects do which makes the generated layout not in compliance with the client needs.

Given that client requirements are mostly communicated through natural language, direct integration into AI-based design processes is difficult. As a result, incorporating NLP into building design systems becomes critical, allowing the translation of complex client language into a format compatible with AI technologies. NLP has found applications in the AEC sector for tasks such as extracting requirements from construction contracts (Hassan & Le, 2020), categorizing requirements in design-build contracts (Ul Hassan *et al.*, 2020), design rule checking (Song *et al.*, 2020), and automating the evaluation of construction specifications using Word2vec word embedding (Moon *et al.*, 2021). Word embeddings serve as a fundamental step in the field of design and architecture, significantly enhancing our ability to understand and process language within the context of various NLP tasks. These potent word representation techniques convert textual information into dense vector spaces, capturing semantic relationships and contextual nuances between words (Ayyadevara, 2018).

The integration of generative AI, particularly ChatGPT, into the architectural design process signifies a transformative shift towards a more efficient and collaborative design workflow. This shift is evident in the use of ChatGPT to diverge ideas during the planning stage of architectural design, proposing a novel concept-making method that combines generative AI with traditional thinking methods, such as the fishbone diagram, to enhance creativity and initial design generation (Maeda *et al.*, 2023). Moreover, the employment of generative AI-powered parametric modeling and BIM further demonstrates the potential of ChatGPT in facilitating rapid exploration of design ideas and producing context-sensitive, creative design generation (Ko *et al.*, 2023). On the other hand, the consideration of client needs in architectural design remains a critical aspect, with methodologies like architectural programming aiming to discover the formal potentialities of sites based on client and expert goals (Al-Shaljia & Al-Dabbaghb, 2022). Such considerations ensure the alignment of architectural designs with client expectations and the fulfillment of user needs, underpinning the essence of a client-centric approach in architectural practice. Collectively, these insights underscore the evolving synergy between AI technologies and architectural design, emphasizing the importance of both innovative AI applications and a deep understanding of client needs in creating responsive and sustainable architectural solutions.

3. Research Methodology

This research develops an AI–BIM RS designed to analyze client preferences and provides personalized recommendations through four steps: data acquisition and processing, BIM database creation, recommendation logic development, and validation. In the data acquisition and processing step, client needs have been collected via survey and subjected to LDA-based topic modeling to identify prevalent client needs. In the BIM database creation step,

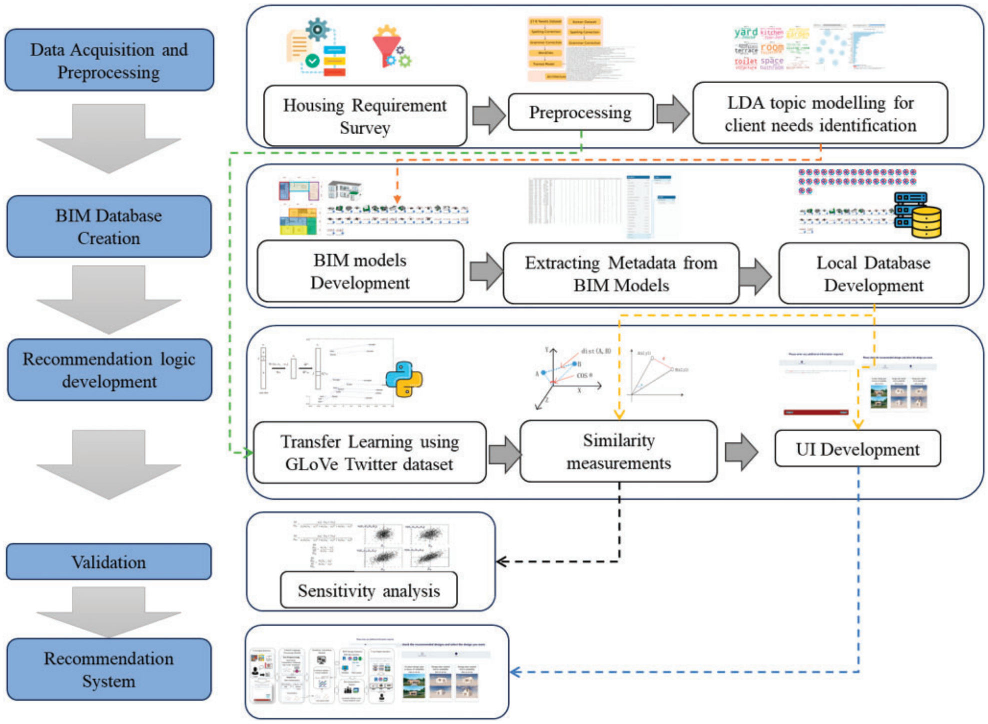

Figure 1: Procedural framework and flowchart.

Table 1: Content of the questionnaire.

Question	Example answers
Please provide your age	20, 30, 40, 50
Specify the intended occupants of the envisioned residence	Individuals living singly, Married individuals, Married individuals with children, Professional
Describe the space configuration of your dream house	3 Bedrooms, 1 guesthouse, 1 living room with floor-to-ceiling glass windows; Bathrooms connected to each room; Kitchen with island; Rooftop patio
Identify 3 to 5 essential features of dream house	1. Rooftop patio 2. Extended room 3. Kitchen island 4. Be sure to include rooms that can be accessed from inside or outside the home 5. There should be a full glass window on one side of the living room

BIM models reflecting these typical needs have been created, and BIM database has been established based on designed schema to store and manage the BIM models. In the recommendation system development step, this research first employs Word2vec modeling to process the identified needs into vectors, which are machine-readable. As core of recommendation system, the recommendation logic has been formulated by leveraging cosine and Euclidean similarity functions to provide recommendations from the BIM database according to vectors of client needs. To ensure an easy-to-use user experience, the user interface (UI) has been developed to enable clients to input their needs for housing and receive corresponding recommendations from the logic. In the validation step, the reliability of the recommendation system has been validated using the sensitivity analysis, focusing on recommendation logic. Figure 1 depicts the overview of research methodology. The details of each step will be discussed in the following subsections.

3.1. Data acquisition and processing

A thorough comprehension of the preferred architectural attributes for homes among different clients underlies this research. To achieve this, our study employed a qualitative survey method to gain an in-depth understanding of the multifaceted preferences, requirements, and aspirations of clients regarding residential spaces. The 586 respondents represent clients and future clients who are interested in building a house in South Korea, ensuring a diverse pool of participants. They answered the open-ended questions on housing features such as type of resident, spatial configuration and layout, and any essential special features as shown in Table 1. The survey was conducted online, making it accessible to anyone interested in home construction or building a house. To enrich the surveyed dataset, web crawling techniques were employed, specifically targeting the housing sector needs of clients. This approach promotes diversity and inclusivity among participants from various geographic locations and socioeconomic backgrounds, thereby ensuring a comprehensive and diverse dataset.

The gathered client responses in Korean from have been examined to identify the most significant patterns and trends in housing needs. To accomplish this, this research has employed a topic modeling process on the survey data. Topic modeling is a natural language processing technique utilized to reveal predominant topics and main themes within text collections. From the array of widely used topic modeling algorithms, we selected LDA, which is specifically designed for textual data, demonstrating its efficacy in discovering hidden themes within extensive text corpora. LDA's ability to handle multidimensional text data could facilitate the identification of key characteristics and essential requirements that are widespread in the realm of housing preferences. For LDA-based topic modeling, this research utilized Gensim library in Python. The pseudo-code 1 for the LDA topic modeling which highlights the preprocessing and visualization techniques is provided below.

Pseudo-code 1: LDA topic modeling of client need surveyed data.

```
Import necessary libraries
Read data from excel file into sentences
Clean and preprocess text in sentences
Function text_cleaning (txt):
  Remove non-ASCII characters, convert to lowercase, replace textual numbers
  Remove special characters and multiple spaces
  Return cleaned text
Apply text_cleaning function to each row in sentences ["content"]
Remove stopwords from cleaned text
Tokenize and lemmatize the text
Generate bigrams and trigrams from lemmatized text
Create a dictionary and doc-term matrix using TF-IDF Model
Train an LDA model with 9 topics
Compute perplexity and coherence score for the LDA model
Visualize topics using WordCloud and pyLDAvis
```

3.2. BIM database creation

In this research, BIM data serve as a digital representation of client housing needs and as a means for providing design option recommendations. In BIM-based projects, architects and designers collaborate to create comprehensive BIM models tailored to specific client requirements. The client needs identified through LDA topic modeling were used as parameters of BIM models. The primary goal of developing BIM models in this study is to encapsulate and represent client-specific housing needs in a digital format. BIM models serve as dynamic and comprehensive information repositories, facilitating the seamless integration of client requirements into the design process. We ensure accuracy and efficiency in collaboratively creating detailed BIM models by using AUTODESK REVIT, thereby improving the overall design process.

For the effective storage and management of BIM model data, this research has established a local database to house BIM models, their rendered pictures, and IFC files. The parameters of the BIM models are represented in a spreadsheet structure. The Excel database structures data into rows and columns, with each row representing an entry and columns defining attributes, such as ID, area total, number of rooms. This database contains 34 entries, and it has been saved as a CSV file for web API use. This file is then read and processed by a Python backend for the web API. This enables us to receive requests and offer recommendations based on the CSV data, responding with suggested BIM models or relevant information as per API request parameters.

A database schema has been logically designed to centralize important BIM model details within the BIM_Models table, ensuring a standardized format while allowing for model-specific flexibility. The computations of the recommendation system are built on top of this schema. The data schema simultaneously stores the metadata that is necessary for design recommendations. This coordinated method makes it possible to retrieve designs quickly and make insightful user recommendations.

3.3. Recommendation logic development

3.3.1. Word2vec *model training*

The choice of WORD2VEC was driven by its balance of simplicity, efficiency, and effective capture of semantic word relationships, fitting our system's initial design requirements. The prevalent WORD2VEC models, CBOW (Continuous Bag of Words) and skip-gram, complement each other, with CBOW capturing broad text topics and skip-gram revealing finer word relationships by predicting target words based on context words. In domain-specific applications with divergent technical vocabularies, pre-trained models struggle due to variation in word frequency distribution and a lack of domain-specific vocabulary. To address this, it is critical to train WORD2VEC with domain-specific knowledge. The refinement process with the GloVe dataset incorporated domain-specific knowledge, including architectural terminology and client preference language, making the model more attuned to the nuances of modular housing design. In this research, transfer learning has been employed to refine a pre-trained model using housing survey data due to the limited size of the collected house specification corpus. Preprocessing the collected dataset to align with the GloVe Twitter dataset's format (Pennington *et al.*, 2014), specifically utilizing its 25-dimensional vector representations, is required for this approach. The GloVe Twitter dataset, derived from 2 billion tweets, provides a nuanced understanding of word semantics in the context of Twitter conversations. The process flow for training the WORD2VEC model and processing the surveyed text is depicted in Fig. 2.

The WORD2VEC model architecture used in this study is depicted in Fig. 3, emphasizing the importance of data transformation. The model uses a fixed number of tokens for each input and includes an input layer with four nodes representing context words. This layer connects to a hidden layer of 100 nodes, which then connects to the output layer via the SoftMax function. The model's simplicity ensures lower computational complexity, while the appropriately sized hidden layer captures word similarity information effectively.

Preprocessing

The Korean dataset preprocessing begins with spell checking using the KoNLPy library. The dataset is then translated into English with the help of Argos-translate, an open-source offline translation library. The decision to translate the dataset from Korean to English prior to applying WORD2VEC, was driven by several considerations. Firstly, the GloVe model, which played a pivotal role in refining our WORD2VEC embeddings, offers a rich, pre-trained dataset based primarily on English language sources. The use of this comprehensive English-based dataset allowed us to leverage a wide array of semantic relationships and nuanced language patterns that are extensively documented and understood within the context of English language research. Moreover, our system aims to serve a broad audience, including stakeholders who might interact with the system in English. Translating the data into English enabled us to create a model that is not only informed by the deep semantic insights available in English NLP resources but is also more accessible to a diverse user base. The pseudo-code 2 for the data preprocessing procedure is provided next.

Pseudo-code 2: Textual data preprocessing.

```
Step 1: Spell Checking and Grammar Correction
function perform_spell_and_grammar_check(dataset):
   corrected_dataset = KoNLPy_spell_check(dataset)
   return corrected_dataset
Step 2: Translation to English
function translate_to_english(dataset):
   translated_dataset = Argos-translate (dataset, source-language =
"Korean," target_language = "English")
   return translated_dataset
Step 3: Main Preprocessing Function
function preprocess_data(raw_dataset):
   Step 1: Spell Checking and Grammar Correction
   dataset_after_spell_check =
perform_spell_and_grammar_check(raw_dataset)
   Step 2: Translation to English
   dataset_in_english = translate_to_english(dataset_after_spell_check)
   return dataset_in_english
```

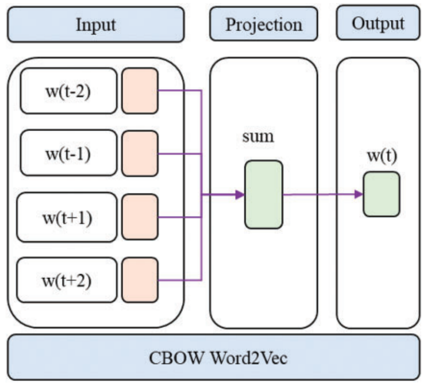

Figure 4: Depiction of token prediction using Word2vec.

Pre-trained model refinement through transfer learning

Word2vec, a three-layer neural network, uses CBOW and skip-gram approaches to transform words into semantic feature vectors (Nicholson, 2023), aiding NLP tasks by capturing contextual relationships. The CBOW architecture (Qiu *et al.*, 2020), depicted with a window size of 4, incorporates context words in the input layer ranging from w(t − 2) to w(t + 2), while capturing semantic links in the hidden layer. Target word prediction in the output layer enables numerical word representations. Figure 4 depicts the core concept of the CBOW Word2vec model used in this study, demonstrating its ability to predict related terms from context words. The mathematical foundations of Word2ve lies in the probability of predicting the target word given context words is calculated by the SoftMax expression (Mikolov *et al.*, (2013).).

The process from translating client input to enhancing our design recommendation system involves a meticulous process where the translated English text, initially segmented into sentences for simplicity, undergoes pre-training with the GloVe model. This step is pivotal as it leverages the rich semantic landscape of the GloVe dataset, renowned for its deep understanding of the English language gleaned from extensive social media data. By arranging the translated sentences in a structured format

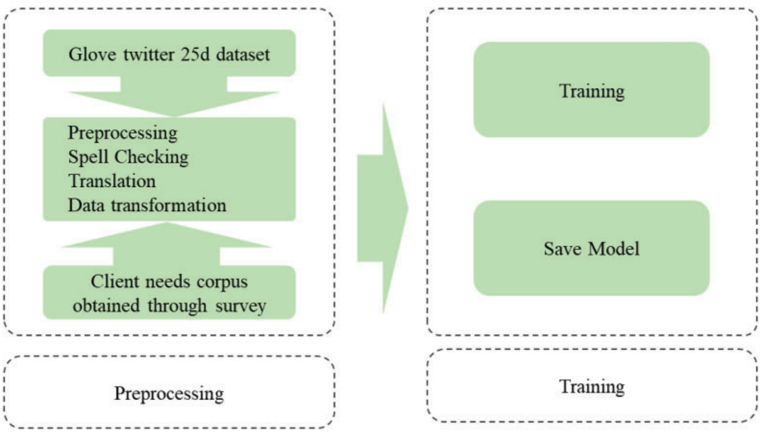

Figure 2: Word2vec training.

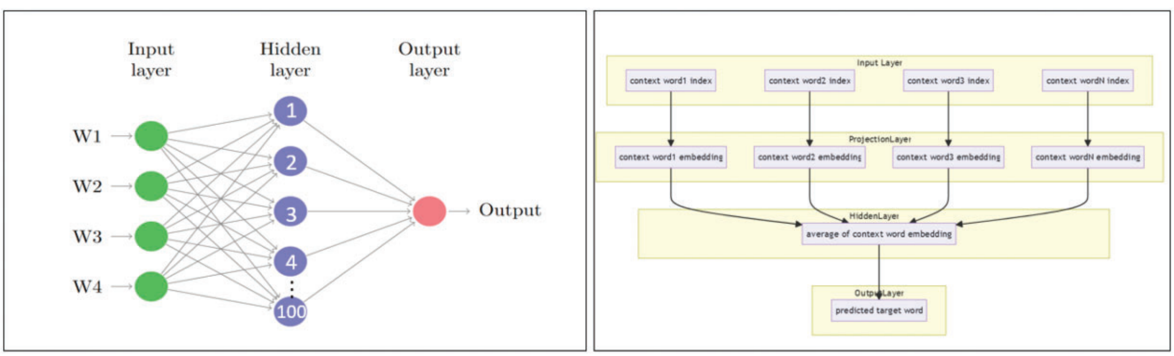

Figure 3: Architecture of Word2vec (modified from Yilmaz & Toklu, 2020).

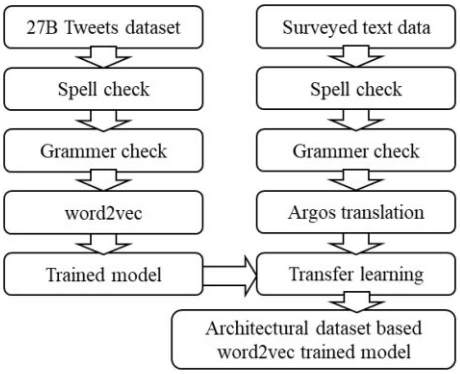

Figure 5: Word2vec trained model training flowchart.

suitable for Word2vec processing, each sentence is methodically fed into the model. Here, Word2vec's neural network architecture springs into action, generating word embeddings that capture the nuanced meanings of words based on their contextual relationships within the corpus. This semantic information, encoded into high-dimensional vectors, is further refined with the GloVe Twitter 25d dataset, enriching our model's grasp of informal and colloquial expressions related to housing. The culmination of this training is a model file in the GloVe format, a compact repository of linguistic knowledge that our system utilizes to interpret client inputs accurately.

The pseudo-code 3 underlying the recommendation process powered by Word2vec-trained model is provided below. This pseudo-code outlines the logical steps involved in analyzing client input, from initial processing to the generation of design recommendations. Furthermore, Fig. 5 visually illustrates the refinement process of our model using the GloVe Twitter 25d dataset, providing a graphical representation of how the pretraining enriches our model's semantic capabilities. Through these methodological steps, our study harnesses NLP techniques to transform client inputs into actionable design insights, paving the way for a more intuitive, responsive, and client-centric architectural design process.

Pseudo-code 3: Word2vec-based recommendations.

```
Prompt the user for input
  Read and store the user input as "user_input"
  Apply NLP processing:
  - Spell correction
  - Grammar correction
  - Argos translation
  - Token finding using predefined criteria
  - If token finding using predefined criteria fails:
    - Token finding using Word2vec synonyms
Recommend the output
  - if valid numeric value is found:
    - Display numeric value in green
  - else:
    - Display error message in red
```

3.3.2. *Similarity measurement*

The use of cosine and Euclidean similarity functions in RS enables effective matching of specific client needs with appropriate building models. The cosine similarity calculates the cosine of the angle between two non-zero vectors, yielding a numeric value that represents the similarity between the client needs and the features of the building models. A closer match is indicated by a higher cosine similarity. Similarly, Euclidean similarity determines the similarity between the client needs and the BIM model features by calculating the straight-line distance between two points in a multidimensional space. Euclidean similarity is effective for comparing quantitative attributes, such as spatial dimensions, offering a straightforward way to find designs closely matching numerical client requirements. Cosine similarity, in contrast, excels in assessing semantic alignment between client preferences and design descriptions, focusing on the directionality of preferences rather than magnitude. This dual approach allows our system to accurately identify designs that are not only close to numerical specifications but also aligned in thematic or conceptual similarity with client requests. The system compares the client input needs (transformed into vectors) with the features of available BIM models to provide recommendations.

In this research, a recommendation logic has been formulated to generate a list of design options based on user input needs, which have been transformed into individual variables using Word2vec. These designs options are identified via database filtering techniques and evaluated through user input. The RS utilized two scoring functions for recommending design choices. Firstly, it uses the Euclidean distance to score, sort the database, and provide the two best design choices. Subsequently, the RS uses cosine similarity to order the database and yield the final design option. These two metrics are employed to mitigate the limitation of scoring function. The Euclidean distance yields a good recommendation when the user input vector is outside the different clusters present in the database, while cosine similarity produces decent results when the user input is inside near to the cluster centroids.

The scoring criterion, which is presented for Euclidean distance, establishes a score normalization procedure that can be used for other metrics. The normalization procedure ensures that the scores from each metric are mutually compatible. This normalization provides an easy metric for judging the resultant design based on user input.

The distance metric computes a distance score for all models in the database based on their distance from user input. Following the score evaluation, the score for each model is normalized.

The normalization maps the score values to a range between 0 and 1, which is displayed in output of recommendation result. Consequently, the top three designs with minimum distance from user requirements are presented as outcome of recommendation logic.

3.3.3. *UI for recommendation system*

The UI of the recommendation needs to empower users to input their requirements through a text form, directing them to design recommendations in accordance with their specified needs. Developed with the Vue.js JavaScript framework, the UI is enriched by CSS and HTML code to imbue the interface with a customized design.

A web based IFC viewer was created for this study in order to visualize the suggested designs. The Three.js package is used for 3D rendering in the IFC viewer, while IFC.js is integrated for IFC file processing. The web application is constructed with a front-end technological stack that consists of HTML, CSS, and JavaScript; JavaScript is used as the main programming language. Through an intuitive web interface, this viewer enables users to interactively explore and evaluate the suggested architectural concepts.

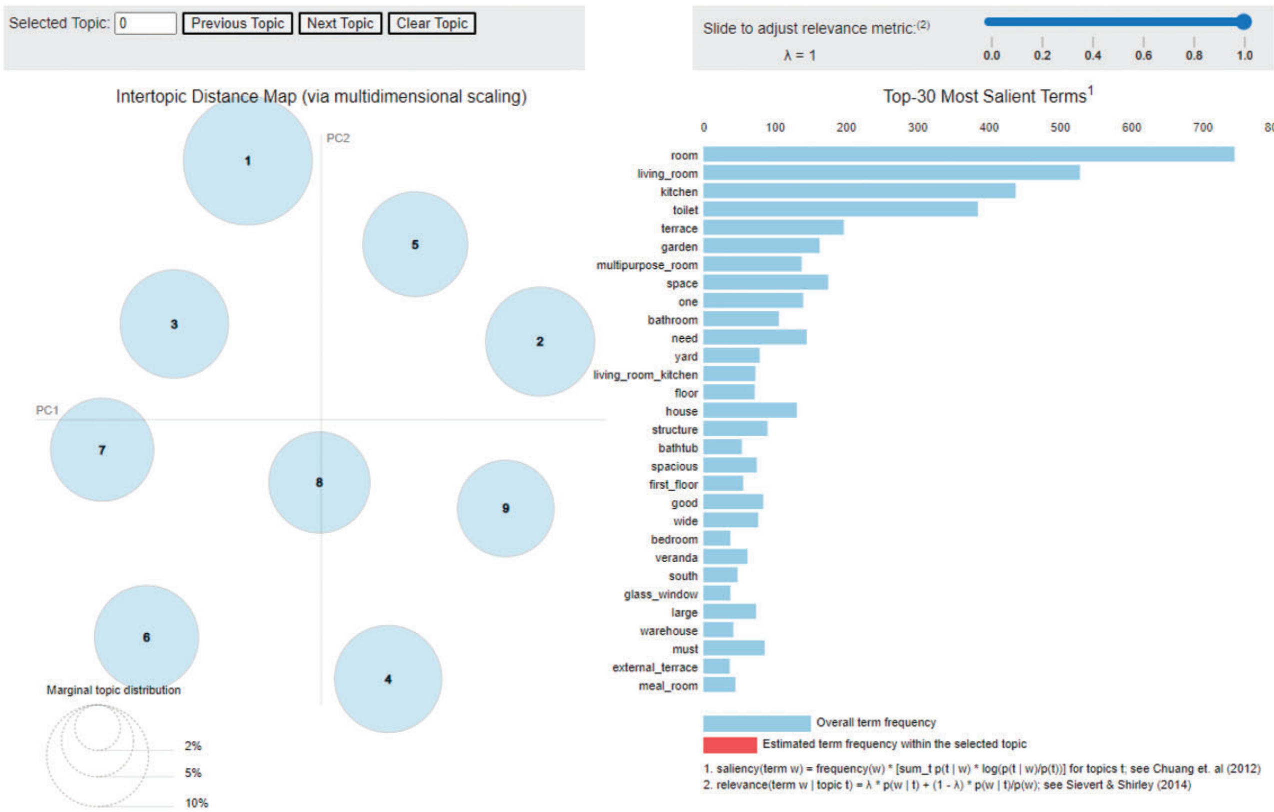

Figure 6: Visual representation of LDA topic modeling.

3.4. Validation

This research employs sensitivity analysis to validate the developed recommendation system, with a specific focus on the weightage scoring metrics (Euclidean and cosine similarity metrics) within recommendation logic. The aim of this analysis is to assess the RS responsiveness to variations in these metrics, ensuring the reliability and robustness of the recommendation algorithms. The weightage scoring metrics undergo fine-tuning and validation using both cosine and Euclidean similarity functions, which are critical for determining the relevance of recommended designs. A sensitivity analysis is conducted to evaluate the extent to which one user input influences the outcome compared with another user input and to evaluate the effectiveness of the employed scoring metrics. To determine the effect of three user inputs on the scoring metric d the effect of each variable on the output can be considered by computing partial derivatives. The effect of user input, u_1,u_2, and u_3 on the recommendation result is given by equations (1) and (2) with respect to u_1,u_2, and u_3.

$$\frac{\frac{\partial d}{\partial u_1}}{\frac{\partial d}{\partial u_2}} = \frac{w_1\,(x_1 - u_1)}{w_2\,(x_2 - u_2)} \tag{1}$$

$$\frac{\frac{\partial d}{\partial u_1}}{\frac{\partial d}{\partial u_3}} = \frac{w_1\,(x_1 - u_1)}{w_3\,(x_3 - u_3)}. \tag{2}$$

Sensitivity analysis entails a systematic adjustment of weightage values to observe their impact on the RS output. This process provides opportunities to identify optimal weightage combinations that resulted in more accurate and relevant suggestions for users by assessing how changes in these metrics influenced the recommendations.

4. Implementation of AI and BIM-based RS

4.1. Client needs for housing buildings

A total of 586 respondents participated in the questionnaire survey, contributing an average of three sentences each, resulting in a total of 1758 sentences directly collected through the questionnaire. Additional 1000 sentences were collected using the keyword “housing” using web crawling. These sentences, derived from forums, articles, and blogs on housing and architectural design, complement the direct survey responses by incorporating broader public discourse on modular housing preferences and trends. The reason for additional textual data was to enhance the dataset’s diversity and comprehensiveness, ensuring it captures a wide spectrum of housing preferences and reflects broader public discourse. This method was employed particularly because the questionnaire survey’s reach might have been potentially limited to participants from only a few nations. By incorporating insights from global online forums, articles, and blogs on housing, we aimed to mitigate geographic biases in the survey responses, thereby enriching our understanding of housing trends across different regions. Following data collection, various cleaning procedures, such as spell-checking and the removal of irrelevant content, were carried out. Following these steps, the dataset was refined to a final set of 2500 sentences for further analysis.

An LDA topic modeling was performed on the surveyed data and the results are shown graphically in Fig. 6. The complete dataset was divided into nine separate subjects for this research, each of which contained a different thematic cluster. A crucial step was determining the optimal number of topics to ensure the model’s effectiveness and accuracy. This decision was based on a comprehensive approach of quantitative evaluation. Initially, we utilized the coherence score, a metric that quantifies the seman-

Figure 7: Word cloud depicting LDA topics.

tic similarity among the top words within each topic, to evaluate the model's performance across a range of topics from 5 to 15. This metric served as a guide to identify the number of topics that maximizes coherence, indicating a more interpretable and meaningful topic structure. Our analysis revealed that a model configured with nine topics achieved a balance, offering high coherence without oversimplifying or excessively fragmenting the themes present in the data.

These nine subjects were further broken down into 30 pertinent terms that stood out clearly in the context of that topic. These carefully chosen terms represent the user demands that have been identified in the field of residential design. These phrases were successfully found because they effectively capture the key characteristics and preferences that direct consumers' desires for residential spaces by embodying a detailed awareness of their requirements.

Figure 7 shows a word cloud depiction of the LDA themes, with each word's size directly corresponding to how frequently it appears in that topic. The most common terms connected to each topic are dynamically captured in this graphical depiction. The predominance of larger-sized words in the analysis indicates the recurring importance of specific features and spaces that clients consistently prioritize in their residential dwellings. This observation emphasizes the critical importance of incorporating identified client requirements into all building BIM models. Incorporating these key elements determined by client preferences ensures that residential designs are aligned with the specific needs and desires of residents, thereby improving overall satisfaction and personalization of living spaces.

Table 2 presents a summary of the consolidated results of LDA analysis, categorizing essential client needs. This compilation serves as a thorough reference, demonstrating the precise mapping between these requirements and relevant BIM model parameters. The table provides a systematic and structured view on the complex relationship between specific client requirements and the elements of BIM.

4.2. BIM database for client needs

To accurately convert client needs into coherent BIMs, the questionnaire's second question plays a crucial role by categorizing standard design layouts into various types according to the occupant type of a house. This differentiation is essential, as it reflects the understanding that houses designed for different occupant types necessitate distinct facilities. For instance, a house intended for a couple with children would include separate rooms for the children, a feature not required in homes for single individuals or couples without children. Leveraging this insight, we meticulously integrated client needs into floor plan layouts tailored to various occupant types. These layouts underwent further revisions by architects and designers to ensure clarity and alignment with client specifications, thereby ensuring BIMs accurately reflect the specific requirements and preferences of its intended occupants.

Using the user needs identified in the housing survey processing, a total of 34 detailed BIM models were created. These models encompassed both single-story and two-story residential configurations, reflecting the survey respondents' diverse architectural preferences. The built models had a variety of floor areas ranging from 15 to 150 m^2, capturing the various spatial requirements articulated by the survey participants. Several types of BIMs were created using Autodesk Revit, covering a wide range of architectural configurations such as different roof types (flat or gable) and floor plan layouts. These models varied in size, taking into account both total area and number of floors, ensuring an accurate representation of various residential architectural styles. A systematic naming convention, as illustrated in Fig. 8, has been established for all BIMs. This convention is grounded in the design concept and the specific properties of the floor plan layout.

Table 3 depicts the mapping between specific client needs and the corresponding BIM parameters for various house building models, providing an overview of how BIM parameters are used to meet various client needs.

Table 2: Client needs and corresponding BIM parameters for house building models.

ID	Client needs	Categories	BIM model parameters	Examples
1	Area	Structural aspects	Floor area	1485.8 m^2
2	Number of floors		Number of floors	2 floors
3	Rooftop		Roof type	Flat roof
4	Roof type		Rooftop area	Rooftop garden
5	South facing house		Orientation (south)	Living spaces facing south direction
6	Living room	Interior design	Living room area	1 living room (27.9 m^2)
7	Attic		Attic space	Converted attic space
8	Veranda		Veranda area	9.3 m^2
9	Kitchen		Kitchen area	1 kitchen (13.9 m^2)
10	Toilet		Number of toilets, toilet area	2 toilets (2.3 m^2 each)
11	Bathroom		Number of bathrooms, bathroom area	2 bathrooms (7.4 m^2 each)
12	Terrace		Terrace area	11.1 m^2
13	Balcony		Balcony area	5.6 m^2
14	Multipurpose room		Multipurpose room area	1 MP room (18.6 m^2)
15	Bedroom		Number of bedrooms, bedroom area	3 bedrooms (18.6 m^2 each)
16	Master bedroom		Number of master bedroom and its area	1 master bedroom (23.2 m^2)
17	Dressing room		Dressing room area	Walk-in closet
18	Library		Library area	1 library (13.9 m^2)
19	Dining room		Dining room area	1 dining room (11.1 m^2)
20	Storage room		Number of storage rooms, storage room area	2 storage rooms (6.7 m^2 each)
21	Kitchen island		Island dimensions, island area	1.2 m x 0.8 m, 0.96 m^2
22	High ceiling	Architectural features	Ceiling height	High ceiling (yes/no) (3.7 m)
23	Glass windows in living room		Number of windows, window height	4 windows panels (3.2 m^2 each)
24	Living room and kitchen together	Floor plan configuration	Open space area (living room & kitchen)	Open concept living and kitchen space
25	Garden/yard		Garden/yard area	Backyard garden with lawn
26	Parking garage		Parking capacity, parking area	Double-car garage, 35.8 m^2

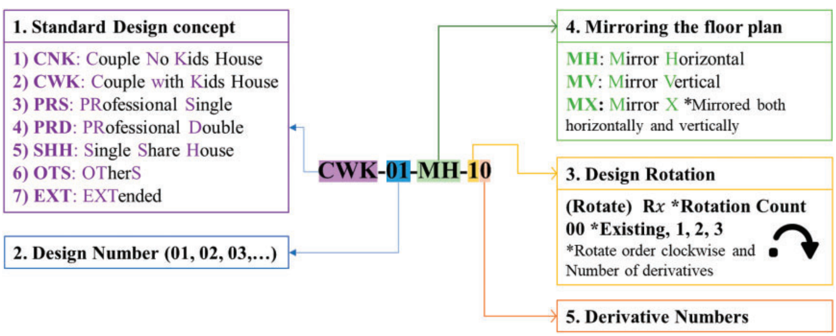

Figure 8: Unique IDs for building models.

Figure 9 illustrates the development of diverse floor plan layouts for BIM models by integrating various client needs into the design.

The level of development (LOD) for the BIM models was meticulously maintained at LOD300, indicating a high level of detailing and accuracy in the geometric and non-geometric information within the model. Additionally, modularity was given top priority when developing these BIM models from conception to completion. The models were simultaneously converted into IFC files, maintaining architectural information for database integration. This two-step technique integrates visual appeal and data-driven functionality, supporting user interaction and architectural understanding in the recommendation system. Figure 10 depicts the floor plan layouts of two distinct BIM designs, emphasizing the adoption of modularity in each layout. The figure also includes the corresponding 3D views of these layouts.

The modularity concept in this research was meticulously developed through a combination of industry collaboration with M3 Systems, a Korean based modular construction company and an in-depth analysis of modular housing production constraints, such as module sizes and panel dimensions, provided by manufacturer from the same company. This approach ensures that the methodology is grounded in practical, real-world considerations, making it applicable and replicable across various contexts within the modular housing industry.

In the implementation of design RS, authors utilized a relational database management system (RDBMS) that is hosted within our local computational environment. This local hosting approach was chosen to facilitate direct control over the database and to simplify the process of integrating the database with our system's architecture. The relational model of the database allows for structured and efficient querying of design options, based on attributes derived from client input and preferences as analyzed by our NLP components. For the purposes of our research and the development of the RS, we employed a RDBMS that could be operated locally. A RDBMS like MySQL serves as a fitting example of

Table 3: Mapping of client needs to building models.

Building model unique ID	ID of client needs included in building model from Table 2
CNK-01-00-00	1, 2, 3, 6, 9, 10, 11, 14, 15, 16, 19, 22, 23, 2
CNK-02-00-00	1, 2, 3, 6, 9, 10, 11, 14, 15, 16, 18, 19, 22, 23, 24
CNK-03-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
CNK-04-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
CNK-05-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
CNK-06-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
CNK-07-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
CNK-08-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
CNK-09-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
CNK-10-00-00	1, 2, 3, 6, 9, 10, 11, 16, 19, 22, 23, 24
CWK-01-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24, 25
CWK-02-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
CWK-03-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
CWK-04-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
CWK-05-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 23, 24, 25
CWK-06-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 23, 24, 25
CWK-07-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
CWK-08-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
CWK-09-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
CWK-10-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
PRO-01-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
PRO-02-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
PRO-03-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24, 25
PRO-04-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24, 25
PRO-05-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
PRO-06-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24, 25
PRO-07-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
PRO-08-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 18, 19, 22, 23, 24, 25
SHH-01-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
SHH-02-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
SHH-03-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
SHH-04-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
SHH-05-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24
SHH-06-00-00	1, 2, 3, 6, 9, 10, 11, 15, 16, 19, 22, 23, 24

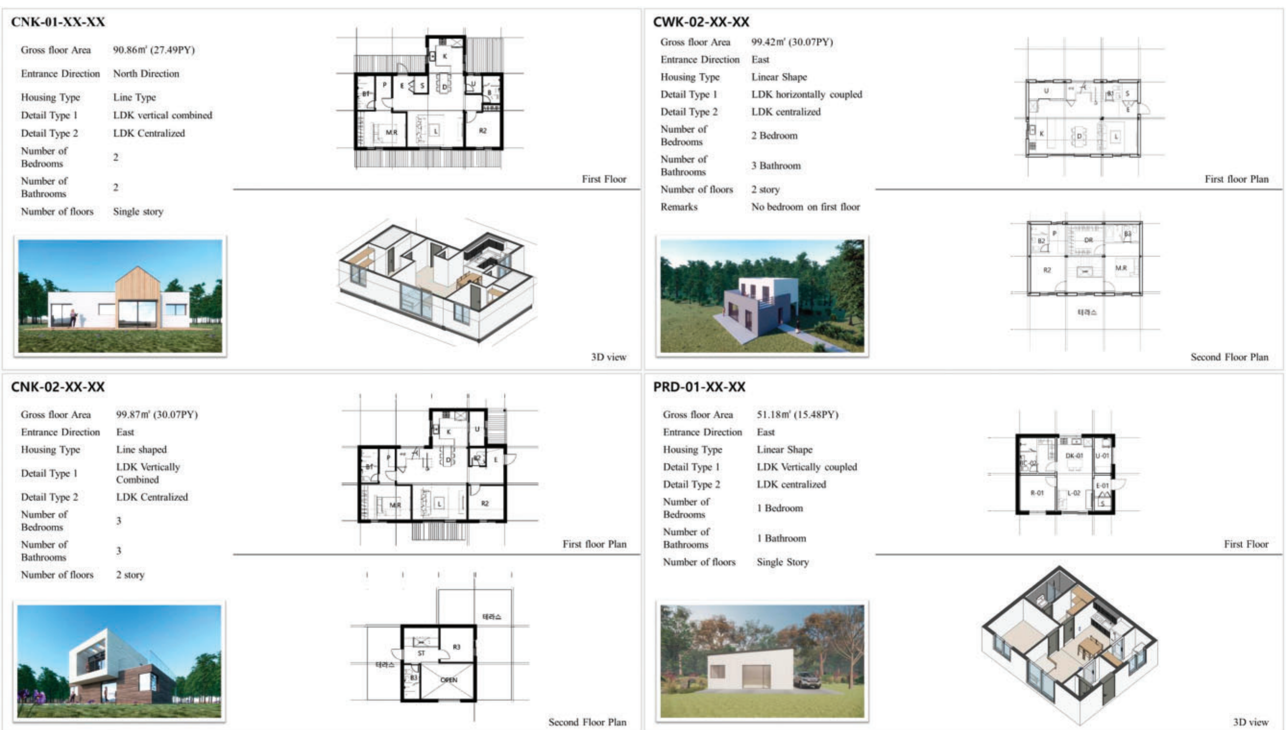

Figure 9: Client needs based floor plan layout design for BIM modeling.

Figure 10: BIM models: Modular floor plans and 3D views combined.

the database technology appropriate for this application, yet the architecture of our system is adaptable, allowing for the selection of particular RDBMS platforms depending on their availability and performance criteria.

Figure 11a and b visually represent the process followed for creating the database and the structured database schema, respectively. The BIM_Models table in this schema contains the essential information about each BIM model, including the number of rooms, bathrooms, floors, and styles. Images and IFC files were separated into individual tables, Images table and IFC_Files table, to maintain data integrity and simplify updates. Clear linkages, established via properties such as "Image_Path" and "IFC_File_Path" ensured seamless connections between tables, facilitating efficient data retrieval, and improving overall information organization and accessibility. Information about the pictures connected to each model is kept on the Images table. The BIM_Models table "Image_Path" property would connect to the relevant entry in this table. Information about the IFC files connected to each model is kept in the IFC_Files Table. Like how photos link to specific records, IFC_File_Path attribute in the BIM_Models table would do the same.

4.3. AI–BIM RS

The AI–BIM RS interprets client input for specific design needs using Word2vec, an NLP technique. The system captures the essence of client requirements by transforming textual descriptions into semantic vectors. These vectors are then compared using similarity functions to identify relevant design elements within the BIM database. Using advanced algorithms, the system suggests design options that closely match the client's input, resulting in tailored and precise solutions. This integration of Word2vec and similarity functions improves the system's understanding of nuanced client preferences, revolutionizing how architects and designers interact with BIM databases to create personalized and client-centered architectural designs.

4.3.1. *System architecture*

AI–BIM RS is the name given to the detached modular house design recommendation system developed in this paper. The system consists of an UI for client requirements input in which the client enters their requirement for house design in the form of text and then submits it. The output page of the UI shows the recommended three modular design alternatives with details such as matching percentages of the client needs and design option and also some parameters of that specific design. The text input of the client is analyzed and read by the Word2vec algorithm and then using cosine and Euclidean similarity recommends the designs. The system architecture of the AI–BIM RS can be seen in Fig. 12.

4.3.2. *Trained* Word2vec *model*

User inputs a text describing their specifications for the house. The Word2vec algorithm contains a list of important architectural keywords or tokens relevant to the recommendation system. These include words like "bedroom," "kitchen", "bathroom", and "balcony". If tokens relevant to the recommendation systems are found in the input text, the recommendation will present a list of three design choices for a house. If the relevant tokens are not found the program will generate semantically similar tokens to the list of tokens present in the program. This step utilizes

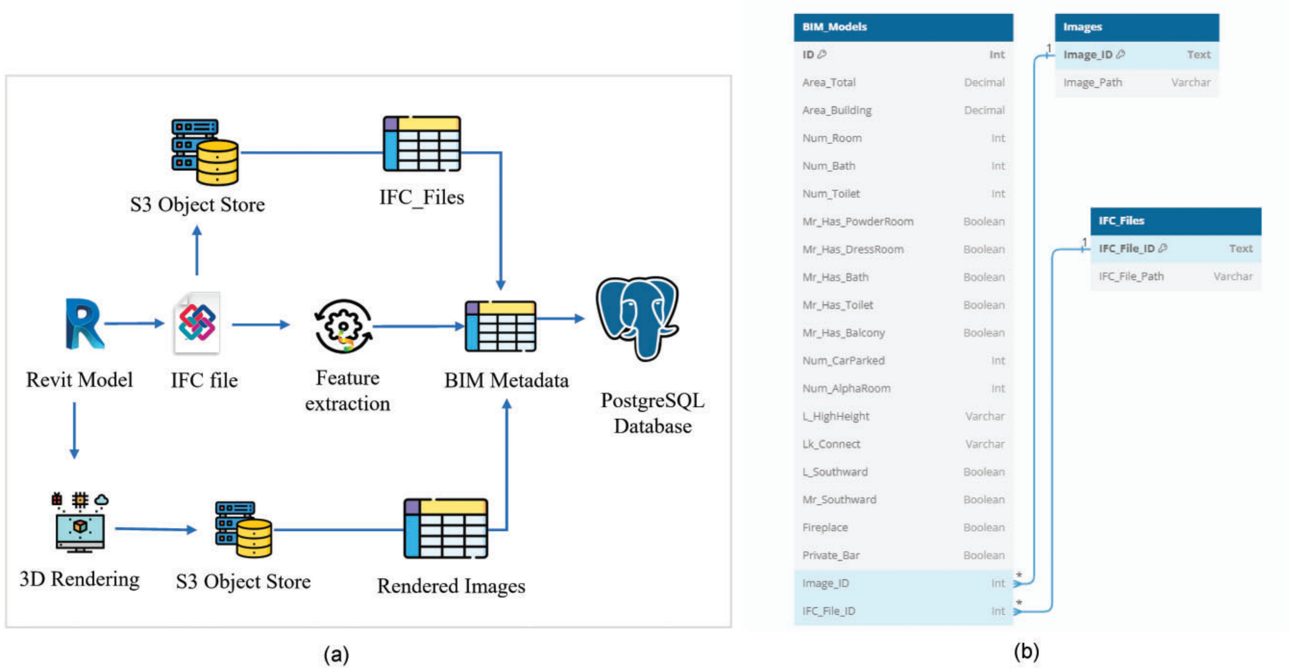

Figure 11: (a) Database creation process. (b) Visual representation of the structured database schema.

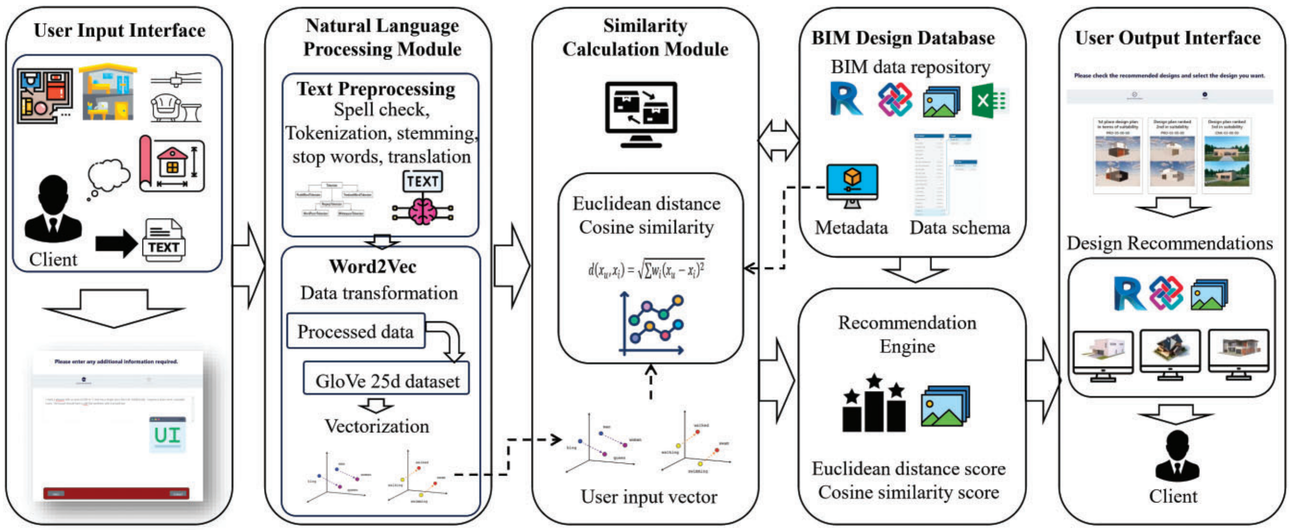

Figure 12: System architecture of the AI–BIM RS.

WORD2VEC model to generate search through the text for newly created tokens as shown in Fig. 13.

While terms like "bedroom," "kitchen," and "bathroom" represent fundamental components of a house, the specificity and thus the real value of client needs emerges from the details surrounding these components. This includes not just the qualitative descriptors of these spaces, such as size or sunlight exposure, but also the quantitative aspects like the number of bedrooms, or specific configurations such as a kitchen with an attached living room.

4.3.3. *Similarity measurement*

The AI–BIM RS employs cosine and Euclidean similarity calculations exclusively between word embedding vectors representing architectural features and client preferences. This approach ensures that the similarity measurements are grounded in semantic relevance, enhancing the system's capability to accurately match client inputs with the most suitable design options from the BIM database.

For instance, when a client specifies a need for "spacious living areas with natural lighting", the system uses WORD2VEC to transform this requirement into a vector representation. This vector is subsequently compared with vectors representing the attributes of available designs in the BIM database. Cosine similarity scores are calculated to identify designs whose features are most semantically aligned with the client's expressed preferences.

Both Euclidean distance and cosine similarity are used to generate a sorted list of suitable design options. Among the three design options, the first two design choices are taken from the top

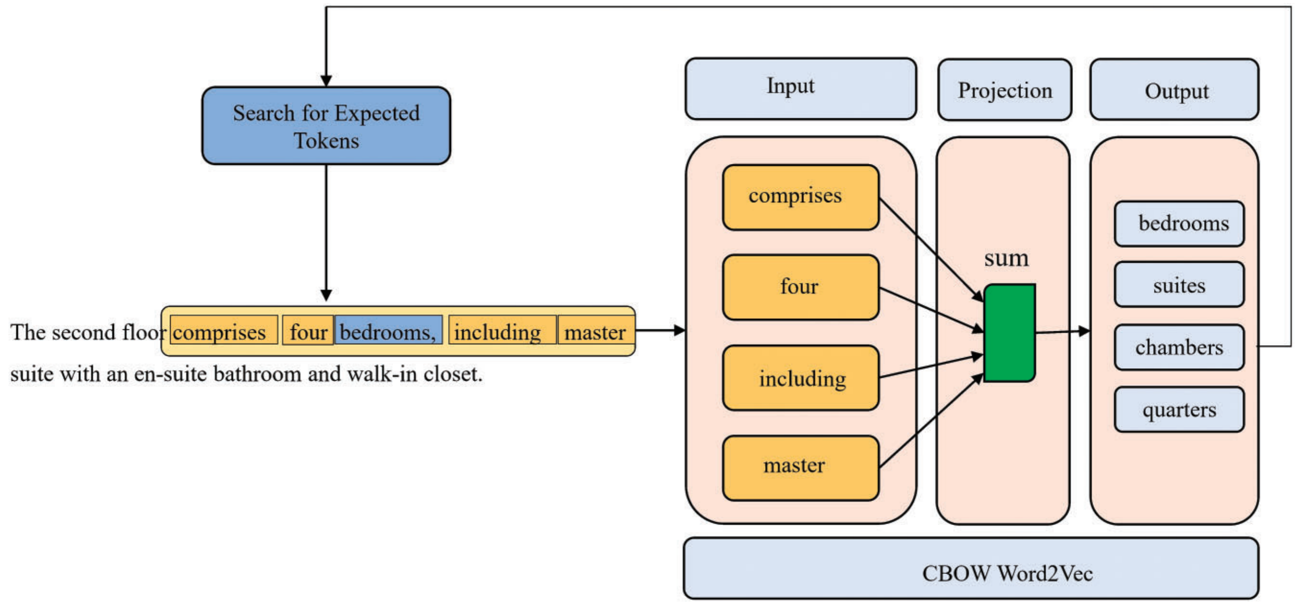

Figure 13: WORD2VEC for similar word prediction.

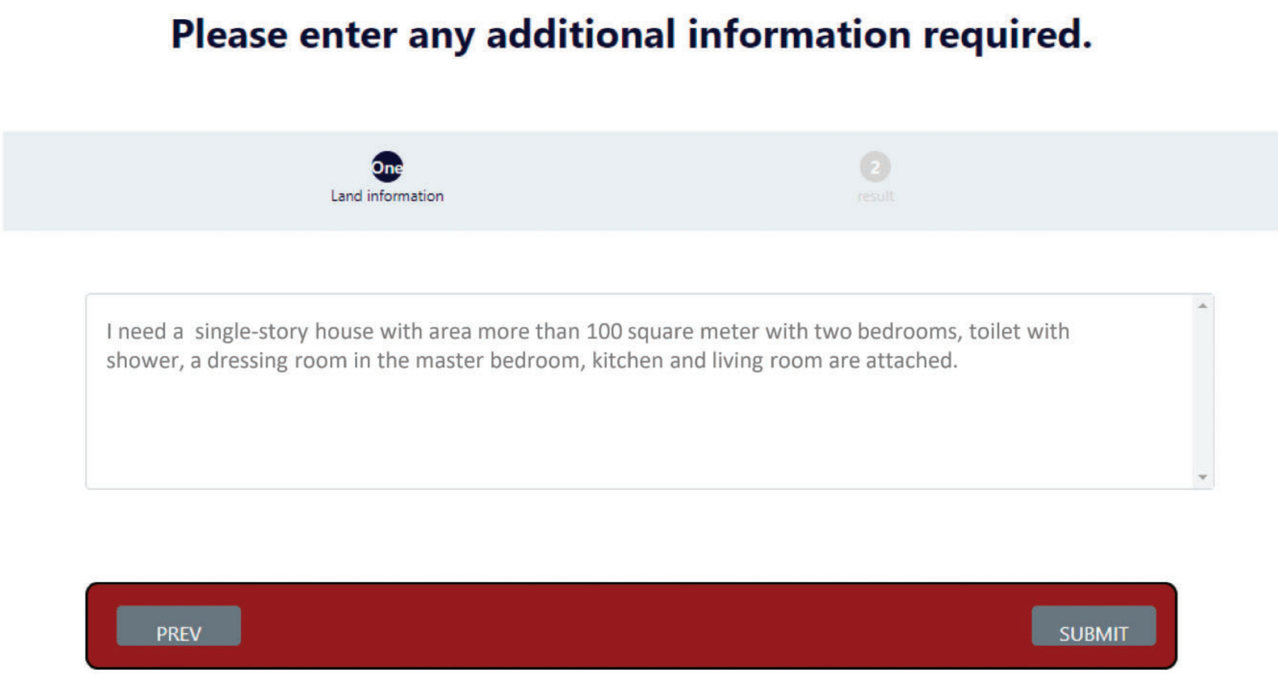

Figure 14: UI for the client requirement input.

two designs from Euclidean scoring whereas the last one is from the best design according to cosine similarity.

4.3.4. *System UI*

The UI includes a user input homepage where the user or client can express their housing needs by typing sentences into a text box. Following submission of the input, the system processes the data and generates three distinct design options. These recommendations, along with details such as unique house IDs and the percentage of match to the user's needs, are displayed on another webpage. Users can choose any model, which is then displayed in a web-based 3D viewer. This viewer uses IFC files to display a 3D representation of the building, allowing users to interactively visualize and explore the recommended designs. Figures 14 and 15 show the client text input and design recommendations output, respectively. The text can be entered in English or Korean language. In the case of Korean, the language is translated in the backend to align it with algorithm.

As depicted in Fig. 15, the system recommends three distinct architectural designs against the client's specified needs. The most suitable design achieved an alignment of 86.70% with the client's requirements, encompassing an area of 109 m^2. To facilitate a deeper exploration of this design, users are provided with a "simulation" button. This feature grants access to a web-based 3D design viewer platform, offering comprehensive details regarding the design, as further illustrated in Fig. 16. This platform enables

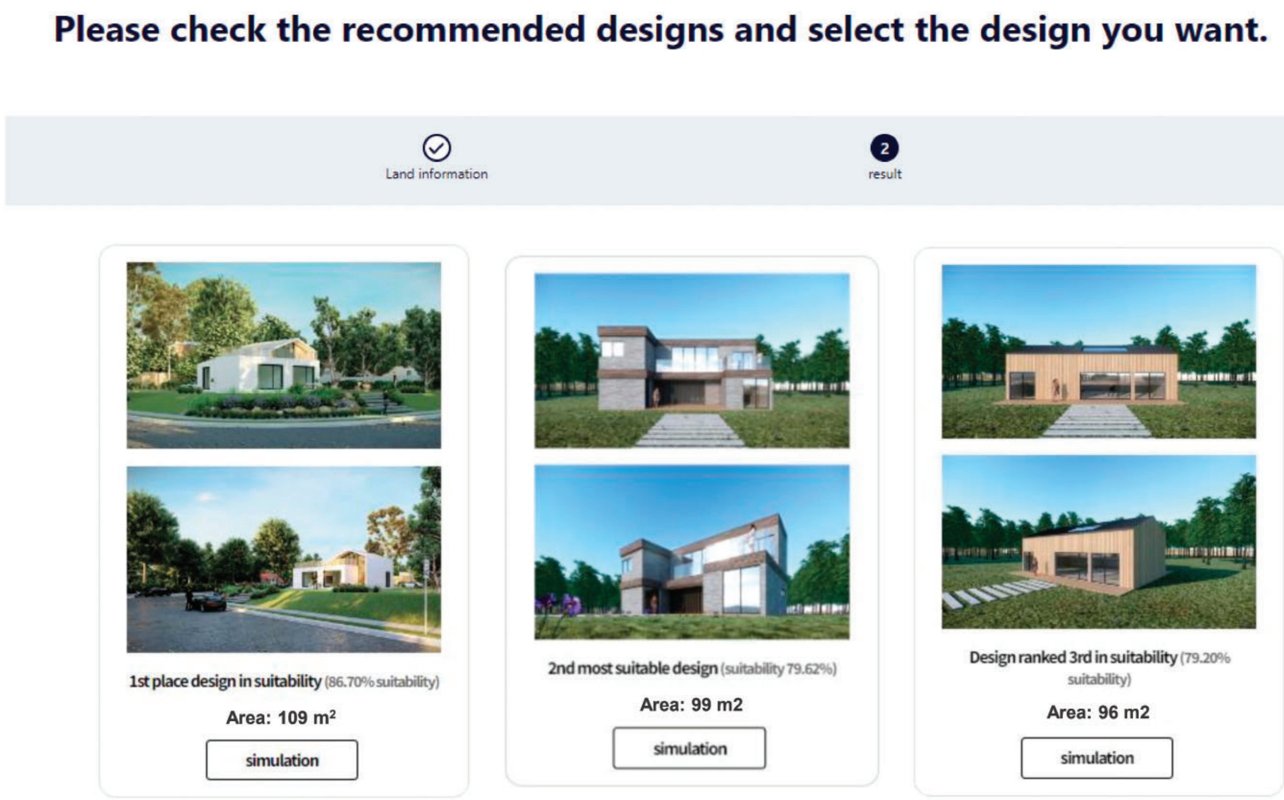

Figure 15: UI for design recommendations result.

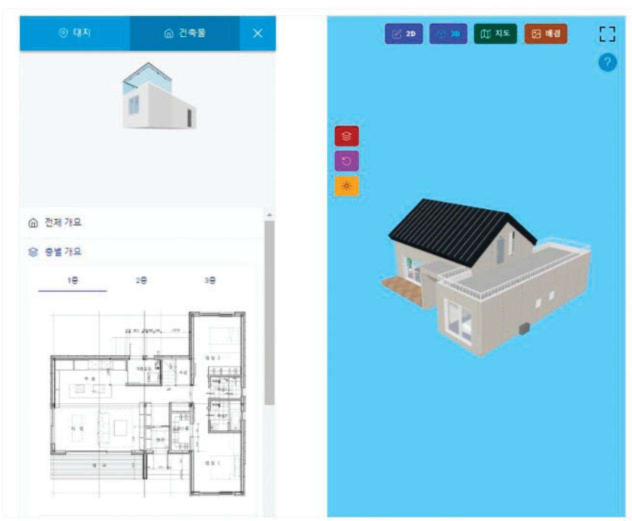

Figure 16: Web based design viewer.

users to thoroughly examine the architectural nuances and specifications of the recommended design.

A detailed visual analysis of Fig. 16 elucidates the layout's components, which comprise two bedrooms, a master bedroom equipped with an en suite toilet and a dressing room, and an interconnected kitchen and living area. This configuration successfully meets all the criteria stipulated in the initial request. Nonetheless, it is important to note a deviation from the client's preference for a single-story layout. The recommended design does not conform to this particular requirement due to the absence of single-story structures within the database that meet the specified layout and space requirements. To address this discrepancy, the recommendation was formulated based on the overall floor plan layout and configuration. The proposed model is inherently modular, allowing for adjustments to better align with client preferences. Specifically, the removal of a module from the attic floor is suggested as a viable solution to reconcile the design with the single-story requirement. This approach underscores the flexibility and adaptability of the recommended design to accommodate specific client needs.

4.4. Validation

The sensitivity analysis performed provides crucial insights into the dynamics of user inputs within the recommendation system. It explains how differences in individual user preferences (u1, u2, and u3) influence the overall recommendation outcome when weighted factors (w1, w2, and w3) are considered as shown in Fig. 17. These figures reveal the complex interplay between user inputs, revealing whether changes in one preference have a greater or lesser influence on the final recommendation than changes in another.

Notably, when w1 = w2 = w3, the analysis reveals a scenario in which all user inputs are given equal weight in determining recommendations. This balanced impact is critical in developing recommendation systems that provide fair and equitable design recommendations, addressing client diverse and evolving needs while remaining responsive to individual preferences.

The line graph in Fig. 18 illustrates the mean squared error (MSE) for various weightage values, providing insights into how changes in weights influence the accuracy of the recommendation system. Each point on the graph represents a different weightage configuration, showcasing the sensitivity of the system to different input weights. The MSE increases with the weightage value.

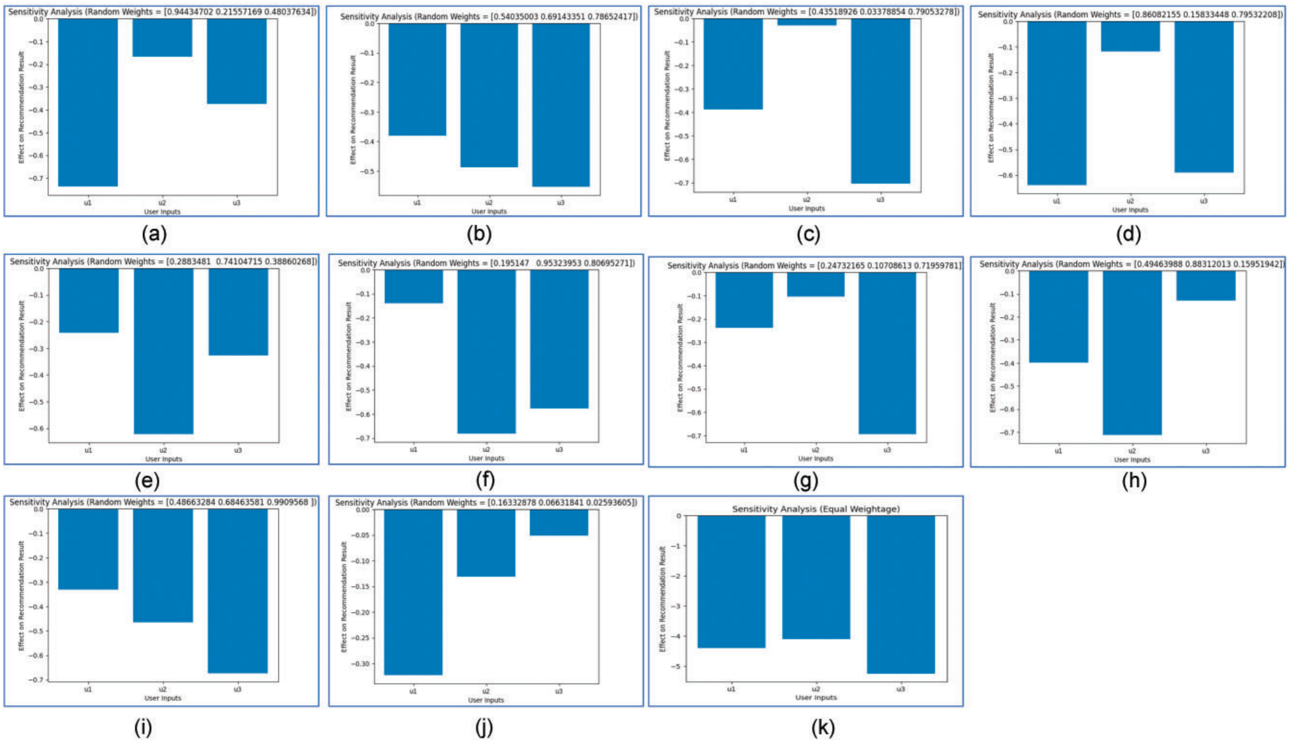

Figure 17: Impact of weightage on recommendation result.

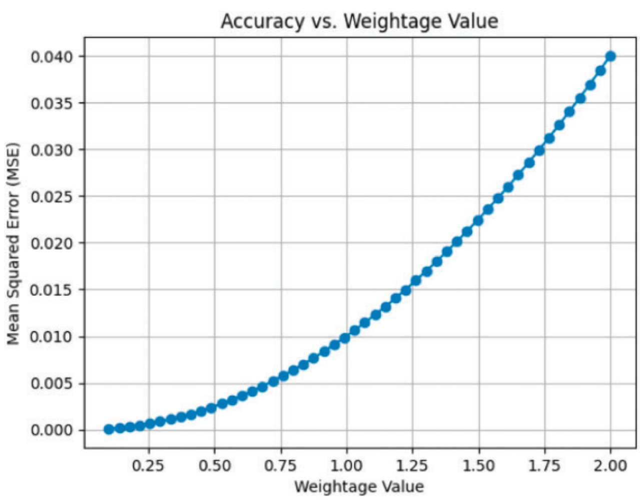

Figure 18: Relationship between weightage value and MSE.

This is because the weightage value determines the importance of each user input. When the weightage value is high, the model will give more importance to that particular user input. This can lead to a higher MSE, as the model may not be able to accurately predict the user's preference. This study gives each user input an identical weight of one because of this. This equal weighting of one is about equivalent to the middle value of all the weightages that were examined.

5. Discussion

A client-needs-based recommendation system was developed using Word2vec and a BIM database, with the goal of improving the architectural design decision-making process. The system generates design options for clients based on textual input and tailors recommendations to their specific needs. The seamless integration of BIM data and Word2vec embeddings demonstrates the significance of combining advanced data analysis techniques with architectural databases. Word2vec's ability to convert textual client needs into numerical vectors enabled addressing the gap between textual input, similarity metrices, and the quantitative nature of BIM parameters, resulting in accurate and relevant recommendations.

The proposed AI–BIM RS is primarily designed for architects, designers, and clients in the modular housing sector, aiming to seamlessly translate client requirements into personalized, feasible modular home designs. Additionally, it serves construction companies and modular home manufacturers by incorporating DfMA principles, enhancing efficiency and client satisfaction in the construction process.

The AEC sector currently lacks a recommendation system which employs NLP to suggest designs based on client preferences from databases. To underscore the effectiveness of AI–BIM RS, it will be critically compared with a collaborative filtering-based RS designed for IoT (internet of things) scenarios developed by Cui Zhihua *et al.* (2020). AI–BIM RS excels in recommendation with its Word2vec model and BIM data integration, providing deep client insights through natural language processing. The collaborative filtering system, on the other hand, relies on clustering and attenuation coefficients, putting nuanced client details at risk. In terms of data utilization, the strength of the AI–BIM RS is its integration of both unstructured (natural language client needs) and structured data (BIM data), allowing for a comprehensive understanding of user requirements and building characteristics. However, the collaborative filtering system focuses on user-item interactions and temporal correlations, which may limit its understanding of the rich context provided by clients. In terms of scalability and efficiency, the scalability of the AI–BIM RS is determined by the efficiency of natural language processing algorithms and BIM data handling. With advancements in these areas, it could handle large-scale datasets efficiently. The collaborative filtering system parallelizes using Spark, demonstrating scalability. However, its effectiveness in diverse and complex IoT scenarios may necessitate further investigation.

The validation process's sensitivity analysis sheds light on the intricate dynamics of user inputs within the recommendation system. It determined how individual user preferences (u1, u2, and u3) influence overall recommendation outcomes and weighted factors (w1, w2, and w3). When $w1 = w2 = w3$, indicating equal weighting of user inputs, the analysis revealed a scenario in which all client needs are given equal weighting. This finding is critical because it indicates that this recommendation system takes a balanced and fair approach by ensuring that the design suggestions are not skewed towards specific preferences by assigning comparable weightage to various client needs, promoting inclusivity and responsiveness to diverse client requirements.

While this system produced promising results, we recognize that it faces some challenges. Handling nuanced and context-specific client needs, such as cultural or environmental preferences, continues to be difficult. Due to a lack of BIM models, the AI–BIM RS encounters a database limitation. A larger database is required to accurately incorporate diverse client needs, ensuring a broader and more precise range of recommendations.

In this research, we leverage the Word2vec approach to capture semantic relationships within textual data, an effective method for understanding and processing natural language in a structured, albeit fixed-dimensional, vector space. This contrasts with the capabilities of the latest generative AI models, such as Large language Model or GPT-3 and beyond, which embody a more dynamic, context-aware processing of language through the utilization of transformers and attention mechanisms. While Word2vec excels in efficiently mapping words to vector spaces based on contextual similarity, providing a solid foundation for specific tasks like similarity searches or basic language understanding, it operates within a narrower scope compared with the expansive, adaptable frameworks offered by generative AI models. These advanced models not only grasp the subtleties of language nuances, idiomatic expressions, and complex grammatical structures through deep learning layers but also generate coherent, contextually relevant text and perform a wide array of language-based tasks with a remarkable degree of fluency and accuracy. The contrast between this research's specific application of NLP in modular housing design and the broader capabilities of generative AI models lies primarily in scope and application focus. This AI–BIM RS emphasizes direct application in architectural design, focusing on translating client requirements into actionable design outputs within a specific domain, whereas generative AI models offer a wide range of functionalities beyond architectural design, including content creation, conversation, and problem-solving across various fields.

While this research utilizes the Word2vec algorithm for semantic understanding and mapping of client requirements to design outcomes, generative AI models like GPT leverage vast amounts of data and sophisticated neural network architectures to understand and generate text. This results in a wide-ranging impact across numerous domains. This paper contributes significantly to the architectural and construction field by offering a specialized application of NLP algorithm, demonstrating the potential for AI to meet specific industry needs, whereas generative AI showcases the expansive and versatile capabilities of AI technology at a broader scale.

Future research could look into incorporating more contextual data into the recommendation process. Furthermore, constant updates and refinement of the Word2vec model are required to adapt to changing language usage and construction trends. In addition, investigating advanced machine learning techniques, such as deep learning models, could improve the system's ability to analyze complex textual data, resulting in more accurate and personalized recommendations. Furthermore, incorporating real-time data sources and user feedback mechanisms would improve the system's adaptability and responsiveness to changing client needs.

6. Conclusion, Limitation, and Future Recommendations

This article endeavors to develop an RS based on NLP that can offer modular housing architectural designs that are customized to meet the needs of the client. In order to accomplish this, a comprehensive client needs for housing survey data was used to gather detailed house features and qualities, which were then incorporated into a BIM model database. Word2vec algorithms were then trained to interpret textual client needs requests and use similarity metrics to generate design suggestions based on the extensive database. To enable smooth interactions between clients and the recommendation system, a carefully designed user-friendly interface was implemented. Furthermore, a comprehensive sensitivity analysis was carried out to improve the weighting of specific client needs, therefore improving the fairness of the comparison framework.

The key findings of this research include the following: first, the development of a modular BIM model database with detailed housing properties and attributes derived from extensive survey data. Second, the training of Word2vec algorithms to analyze client needs expressed in text and use similarity metrics for design recommendations which provides an approach of how these can be integrated with BIM database. Third, a thorough sensitivity analysis was conducted to optimize the allocation of weights to individual client needs, refining the comparison framework in an equitable manner.

This study makes several significant contributions to the AEC sector as follows: firstly, it introduces an innovative detached housing recommendation system powered by NLP techniques which effectively bridges the gap between client needs and tailored design recommendations, improving the overall user experience. Secondly, the system's ability to process client-input design requirements in textual format using Word2vec embeddings represents a significant advancement. Thirdly, the study improves client-architect communication by providing a user-friendly interface that encourages engagement. Fourthly, the integration of BIM data and Word2vec embeddings highlights the value of combining advanced data analysis techniques with architectural databases. The ability of Word2vec to convert textual client needs into numerical vectors allowed us to bridge the gap between textual input and the quantitative nature of BIM parameters, resulting in accurate and relevant recommendations. Finally, the research provides important data for Architect's future design decisions by providing valuable insights into client preferences and emerging design trends.

In recognizing the vital role that industrial producers play in the modular housing industry; we acknowledge the necessity of integrating their technical limitations and expertise into the generative process of BIM models. The collaboration with industrial producers not only ensures that the design recommendations are feasible and aligned with current manufacturing capabilities but also enhances the practicality and applicability of the proposed AI–BIM RS. Inclusion of industry-specific constraints and production insights could significantly refine the system's outputs, ensuring that the recommended designs are not only client-centered but also technically viable and cost-effective. Future iterations of the system will aim to incorporate a feedback loop with industrial producers to regularly update the BIM model database with constraints and innovations from the production side. This integration promises to bridge the gap between client desires and manufacturing realities, fostering a more efficient, innovative, and collaborative ecosystem in modular housing design.

Besides many significances there are still some limitations to this RS which needs further improvements. One of the limitations is the size of the database, which has become quite substantial, necessitating a corresponding increase in its scale. Furthermore, certain client needs, particularly those related to environmental sustainability and cultural considerations for housing, are not adequately represented in the database. Additionally, the system lacks detailed information on external building features, such as the absence of data pertaining to amenities like garages.

To address these limitations, future research endeavors could incorporate innovative methods and approaches. Firstly, the adoption of generative design techniques such as GANs or Genetic Algorithms could be explored to create custom BIM models from the existing database, potentially mitigating the need for an excessively large database. This strategy may streamline the system and enhance efficiency. Secondly, addressing the absence of client needs related to environmental sustainability could involve conducting another survey specifically focused on these aspects, enabling the integration of eco-friendly features into the housing recommendations. Finally, to enrich the database with details on external building features, such as garages or swimming pools, ongoing efforts should include systematically updating the models with additional attributes, ensuring a more comprehensive representation of diverse client preferences and requirements. In addition, future research could integrate advanced language models like GPT and ChatGPT to enhance semantic understanding and personalization in architectural design RSs, leveraging their ability to interpret complex client requests and facilitate interactive user experiences.

Acknowledgments

This work is supported in 2024 by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269).

Conflict of interest statement

None declared.

References

Abanda, F. H., Tah, J. H. M., & Cheung, F. K. T. (2017). BIM in off-site manufacturing for buildings. *Journal of Building Engineering*, **14**, 89–102. https://doi.org/10.1016/j.jobe.2017.10.002.

Abd Razak, M. I., Khoiry, M. A., Wan Badaruzzaman, W. H., & Hussain, A. H. (2022). DfMA for a better industrialised building system. *Buildings*, **12**(6), 794. https://doi.org/10.3390/buildings12060794.

Abdul Nabi, M., & El-adaway, I. H. (2020). Modular construction: Determining decision-making factors and future research needs. *Journal of Management in Engineering*, **36**(6), 04020085. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000859.

Adeli, H. (2001). Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, **16**, 126–142. https://doi.org/10.1111/0885-9507.00219.

Al-Shaljia, F. F., & Al-Dabbaghb, A. H. (2022). Employing architectural programming to discover the formal potentialities of the site issue for the reconstruction of old Mosul. *Muthanna Journal of Engineering and Technology*, **10**(1), 9–16. https://doi.org/10.52113/3/eng/mjet/2022-10-01/09-16.

Almashaqbeh, M., & El-Rayes, K. (2021). Optimizing the modularization of floor plans in modular construction projects. *Journal of Building Engineering*, **39**, 102316. https://doi.org/10.1016/j.jobe.2021.102316.

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: An overview. *Journal of Physics: Conference Series*, **1142**, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012.

Amara, S., & Subramanian, R. R. (2020). Collaborating personalized recommender system and content-based recommender system using TextCorpus. In *Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 105–109). IEEE. https://doi.org/10.1109/ICACCS48705.2020.9074360.

Architectural Designs. (2023). Selling quality house plans for generations. Architectural Designs. https://www.architecturaldesigns.com/ Accessed 10 October 2023.

Ayyadevara, V. K. (2018). Word2vec. In *Pro machine learning algorithms: A hands-on approach to implementing algorithms in Python and R* (pp. 167–178). Apress. https://doi.org/10.1007/978-1-4842-3564-5_8.

Bakhshi, S., Chenaghlou, M. R., Rahimian, F. P., Edwards, D. J., & Dawood, N. (2022). Integrated BIM and DfMA parametric and algorithmic design-based collaboration for supporting client engagement within offsite construction. *Automation in Construction*, **133**, 104015. https://doi.org/10.1016/j.autcon.2021.104015.

Bao, Z., Laovisutthichai, V., Tan, T., Wang, Q., & Lu, W. (2022). Design for manufacture and assembly (DfMA) enablers for offsite interior design and construction. *Building Research & Information*, **50**(3), 325–338. https://doi.org/10.1080/09613218.2021.1966734.

Bertram, N., Fuchs, S., Mischke, J., Palter, R., Strube, G., & Woetzel, J. (2019). Modular construction: From projects to products. *McKinsey & Company: Capital Projects & Infrastructure*, **1**, 1–34. https://www.mckinsey.com/capabilities/operations/our-insights/modular-construction-from-projects-to-products.

Bianconi, F., Filippucci, M., & Buffi, A. (2019). Automated design and modeling for mass-customized housing. A web-based design space catalog for timber structures. *Automation in Construction*, **103**, 13–25. https://doi.org/10.1016/j.autcon.2019.03.002.

Brandão, D. Q. (2011). Technical recommendations and guidelines for designing adaptable low-income houses. *Ambiente Construído*, **11**, 73–96. https://doi.org/10.1590/S1678-86212011000200006.

Cui, Z., Xu, X., Fei, X. U. E., Cai, X., Cao, Y., Zhang, W., & Chen, J. (2020). Personalized recommendation system based on collaborative filtering for IoT scenarios. *IEEE Transactions on Services Computing*, **13**(4), 685–695. https://doi.org/10.1109/TSC.2020.2964552.

Dauphin Americas. (2022). The power of BIM. Using BIM in the design process means more satisfied customers, reduced costs, safer working environments, and more. https://blog.dauphin.com/en/the-power-of-bim Accessed 17 October 2023.

Designing Buildings. (2022). Client requirements. https://www.designingbuildings.co.uk/wiki/Client_requirements Accessed 5 October 2023.

El Mounla, K., Beladjine, D., Beddiar, K., & Mazari, B. (2023). Lean-BIM approach for improving the performance of a construction project in the design phase. *Buildings*, **13**(3), 654. https://doi.org/10.3390/buildings13030654.

Gan, V. J. (2022). BIM-based graph data model for automatic generative design of modular buildings. *Automation in Construction*, **134**, 104062. https://doi.org/10.1016/j.autcon.2021.104062.

Gao, X., Asami, Y., Zhou, Y., & Ishikawa, T. (2013). Preferences for floor plans of medium-sized apartments: A survey analysis in Beijing. *China Housing Studies*, **28**(3), 429–452. https://doi.org/10.1080/02673037.2013.759542.

Gbadamosi, A. Q., Mahamadu, A. M., Oyedele, L. O., Akinade, O. O., Manu, P., Mahdjoubi, L., & Aigbavboa, C. (2019). Offsite construction: Developing a BIM-based optimizer for assembly. *Journal of Cleaner Production*, **215**, 1180–1190. https://doi.org/10.1016/j.jclepro.2019.01.113.

Ghannad, P., & Lee, Y. C. (2022). Automated modular housing design using a module configuration algorithm and a coupled generative adversarial network (CoGAN). *Automation in Construction*, **139**, 104234. https://doi.org/10.1016/j.autcon.2022.104234.

Hassan, F. U., & Le, T. (2020). Automated requirements identification from construction contract documents using natural language processing. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, **12**(2), 04520009. https://doi.org/10.1061/(ASCE)LA.1943-4170.0000379.

Hentschke, C. S., Formoso, C. T., Rocha, C. G., & Echeveste, M. E. (2014). A method for proposing valued-adding attributes in customized housing. *Sustainability*, **6**(12), 9244–9267. https://doi.org/10.3390/su6129244.

Heravi, G., & Eslamdoost, E. (2015). Applying artificial neural networks for measuring and predicting construction-labor productivity. *Journal of Construction Engineering and Management*, **141**, 04015032. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001006.

House Plans. (2023). Welcome to houseplans! Find your dream home today! Search from nearly 40,000 plans. Houseplans.com. https://www.houseplans.com/ Accessed 10 October 2023.

Hwang, K. E., & Kim, I. (2022). Post-COVID-19 modular building review on problem-seeking framework: Function, form, economy, and time. *Journal of Computational Design and Engineering*, **9**(4), 1369–1387. https://doi.org/10.1093/jcde/qwac057.

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods, and evaluation. *Egyptian Informatics Journal*, **16**(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005.

Kamali, M., & Hewage, K. (2016). Life cycle performance of modular buildings: A critical review. *Renewable and Sustainable Energy Reviews*, **62**, 1171–1183. https://doi.org/10.1016/j.rser.2016.05.031.

Karan, E., & Asadi, S. (2019). Intelligent designer: A computational approach to automating design of windows in buildings. *Automation in Construction*, **102**, 160–169. https://doi.org/10.1016/j.autcon.2019.02.019.

Ko, J., Ajibefun, J., & Yan, W. (2023). Experiments on Generative AI-powered parametric modeling and BIM for architectural design. preprint (arXiv:2308.00227). https://doi.org/10.48550/arXiv.2308.00227.

Lawson, R. M., Ogden, R. G., & Bergin, R. (2012). Application of modular construction in high-rise buildings. *Journal of Architectural Engineering*, **18**(2), 148–154. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000057.

Lee, Y., Kim, J. I., Flager, F., & Fischer, M. (2022). Methodology to estimate logistics costs for vertically transported prefabricated wall panels. *Journal of Computational Design and Engineering*, **9**(4), 1348–1368. https://doi.org/10.1093/jcde/qwac055.

Li, Y., Wei, H., Han, Z., Huang, J., & Wang, W. (2020). Deep learning-based safety helmet detection in engineering management based on convolutional neural networks. *Advances in Civil Engineering*, **2020**, 9703560. https://doi.org/10.1155/2020/9703560.

Lu, N., & Korman, T. (2010). Implementation of building information modeling (BIM) in modular construction: Benefits and challenges. In *Proceedings of the Construction Research Congress 2010: Innovation for Reshaping Construction Practice*(pp. 1136–1145). https://doi.org/10.1061/41109(373)114.

Maeda, Y., Ito, J., & Kado, K. (2023). Design process with generative AI and thinking methods: Divergence of ideas using the fishbone diagram method. In Karwowski W., & Ahram T. (Eds.), *Artificial intelligence, social computing and wearable technologies*. https://doi.org/10.54941/ahfe1004191.

Merrell, P., Schkufza, E., & Koltun, V. (2010). Computer-generated residential building layouts. In *Proceedings of the ACM SIGGRAPH Asia 2010 Papers (SIGGRAPH ASIA '10)*(pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/1866158.1866203.

,Mikolov, T., ,Sutskever, I., ,Chen, K., & ,Corrado, G. S.,(2013).Distributed representations of words and phrases and their compositionality. Proceeding of Advances in Neural Information Processing Systems 26 (NIPS 2013)

Modular Building Institute. (2021). *Permanent modular construction report*. https://www.modular.org/industry-analysis/ Accessed 2 February 2023.

Moon, S., Lee, G., Chi, S., & Oh, H. (2021). Automated construction specification review with named entity recognition using natural language processing. *Journal of Construction Engineering and Management*, **147**(1), 04020147. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001953.

National Institute of Building Science. (2018). Report of the result of the 2018 off-site construction industry survey. https://www.nibs.org/files/pdfs/NIBS_OSCC_SurveyReport_2018.pdf Accessed 2 February 2023.

Nguyen, B. N., London, K., & Zhang, P. (2022). Stakeholder relationships in off-site construction: A systematic literature review. *Smart and Sustainable Built Environment*, **11**(3), 765–791. https://doi.org/10.1108/SASBE-11-2020-0169.

Nicholson, C. V. (2023). A beginner's guide to Word2vec and neural word embeddings. Pathmind. https://wiki.pathmind.com/word2vec Accessed 10 October 2023.

Othman, A. A. E. (2015). An international index for customer satisfaction in the construction industry. *International Journal of Construction Management*, **15**(1), 33–58. https://doi.org/10.1080/15623599.2015.1012140.

Pan, W., & Sidwell, R. (2011). Demystifying the cost barriers to offsite construction in the UK. *Construction. Management and Economics*, **29**(11), 1081–1099. https://doi.org/10.1080/01446193.2011.637938.

Park, D. Y., Choi, J., Ryu, S., & Kim, M. J. (2022). A user-centered approach to the application of BIM in smart working environments. *Sensors*, **22**(8), 2871. https://doi.org/10.3390/s22082871.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*(pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162.

Qiu, D., Jiang, H., & Chen, S. (2020). Fuzzy information retrieval based on continuous bag-of-words model. *Symmetry*, **12**(2), 225. https://doi.org/10.3390/sym12020225.

Rankohi, S., Bourgault, M., Iordanova, I., & Carbone, C. (2023). Developing a construction oriented DfMA deployment framework. *Buildings*, **13**(4), 1050. https://doi.org/10.3390/buildings13041050.

Rehman, S. U., Kim, I., & Choi, J. (2023). Data-driven integration framework for four-dimensional building information modeling simulation in modular construction: A case study approach. *Journal of Computational Design and Engineering*, **10**(6), 2288–2311. https://doi.org/10.1093/jcde/qwad100.

Rehman, S. U., Ryu, S., & Kim, I. (2022). An analysis and consolidation of DfMA based construction guidelines and its validation through a Korean case study. In *Proceedings of the International Conference on Geometry and Graphics*(pp. 749–759). Springer International Publishing. https://doi.org/10.1007/978-3-031-13588-0_65.

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, **9**(1), 59. https://doi.org/10.1186/s40537-022-00592-5.

Sebastian, R. (2011). Changing roles of the clients, architects, and contractors through BIM. *Engineering, Construction, and Architectural Management*, **18**(2), 176–187. https://doi.org/10.1108/09699981111111148.

Sebastian, R., Haak, W., & Vos, E. (2009). BIM application for integrated design and engineering in small-scale housing development: A pilot project in the Netherlands. In *Proceedings of the International Symposium CIB-W096 Future Trends in Architectural Management*(pp. 2–3). https://www.irbnet.de/daten/iconda/CIB14394.pdf.

Shin, J., Rajabifard, A., Kalantari, M., & Atazadeh, B. (2020). Applying BIM to support dispute avoidance in managing multi-owned buildings. *Journal of Computational Design and Engineering*, **7**(6), 788–802. https://doi.org/10.1093/jcde/qwaa057.

Song, J., Lee, J. K., Choi, J., & Kim, I. (2020). Deep learning-based extraction of predicate-argument structure (PAS) in building design rule sentences. *Journal of Computational Design and Engineering*, **7**(5), 563–576. https://doi.org/10.1093/jcde/qwaa046.

Spence, C. (2020). Senses of place: Architectural design for the multisensory mind. *Cognitive Research: Principles and Implications*, **5**(1), 46. https://doi.org/10.1186/s41235-020-00243-4.

Stem, J. M. A. (n.d.). *House plans, garage plans, duplex plans, & floor plans at family home plans. Family home plans*. https://www.familyhomeplans.com/ Accessed 10 October 10, 2023.

Svensson, C., & Barfod, A. (2002). Limits and opportunities in mass customization for "build to order" SMEs. *Computers in Industry*, **49**(1), 77–89. https://doi.org/10.1016/S0166-3615(02)00060-X.

Ul Hassan, F., Le, T., & Tran, D. H. (2020). Multi-class categorization of design-build contract requirements using text mining and natural language processing techniques. In *Proceedings of the Construction Research Congress 2020*(pp. 1266–1274). American Society of Civil Engineers. https://doi.org/10.1061/9780784482889.135.

Velamati, S. (2012). *Feasibility, benefits, and challenges of modular construction in high rise development in the United States: A developer's perspective*. Doctoral dissertation, Massachusetts Institute of Technology. https://hdl.handle.net/1721.1/77129.

Wei, Y. Z., Moreau, L., & Jennings, N. R. (2005). Learning users' interests by quality classification in market-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, **17**(12), 1678–1688. https://doi.org/10.1109/TKDE.2005.200.

Westbrook, R., & Williamson, P. (1993). Mass customization: Japan's new frontier. *European Management Journal*, **11**(1), 38–45. https://doi.org/10.1016/0263-2373(93)90022-A.

Wilmot, C. G., & Mei, B. (2005). Neural network modeling of highway construction costs. *Journal of Construction Engineering and Management*, **131**, 765–771. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:7(765).

Wilson, J. (2019). Design for modular construction: An introduction for architects. *Modular Advantage Magazine. American Institute of Architects, National Institute of Building Sciences*, **41**. https://offsiteconstructionnetwork.com/design-for-modular-construction-an-introduction-for-architects .

Wuni, I. Y., & Shen, G. Q. (2020). Barriers to the adoption of modular integrated construction: Systematic review and meta-analysis, integrated conceptual framework, and strategies. *Journal of Cleaner Production*, **249**, 119347. https://doi.org/10.1016/j.jclepro.2019.119347.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, **39**(1), 93–112. https://doi.org/10.1177/0739456X17723971.

Yap, J. B. H., & Skitmore, M. (2018). Investigating design changes in Malaysian building projects. *Architectural Engineering and Design Management*, **14**(3), 218–238. https://doi.org/10.1080/17452007.2017.1384714.

Yilmaz, S., & Toklu, S. (2020). A deep learning analysis on question classification task using Word2vec representations. *Neural Computing and Applications*, **32**(7), 2909–2928. https://doi.org/10.1007/s00521-020-04725-w.

Yu, Y., Ha, D., Lee, K., Choi, J., & Koo, B. (2022). ArchShapesNet: A novel dataset for benchmarking architectural building information modeling element classification algorithms. *Journal of Computational Design and Engineering*, **9**(4), 1449–1466. https://doi.org/10.1093/jcde/qwac064.

Zawidzki, M., & Szklarski, J. (2020). Multi-objective optimization of the floor plan of a single-story family house considering position and orientation. *Advances in Engineering Software*, **141**, 102766. https://doi.org/10.1016/j.advengsoft.2019.102766.

Received: January 21, 2024. **Revised:** April 15, 2024. **Accepted:** April 15, 2024