T.C.

REPUBLIC OF TURKEY
AYDIN ADNAN MENDERES UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
CIVIL ENGINEERING

A NOVEL TREND PREDICTION SYSTEM DESIGN VIA ASSOCIATION RULE MINING REPRESENTED IN KNOWLEDGE GRAPHS: BIM RESEARCH FIELD APPLICATION

GAMZE GÜLSÜN GAZİ

MASTER'S THESIS

Supervisors:
Assoc. Prof. Dr. Gözde Başak ÖZTÜRK
Asst. Prof. Dr. Fatih SOYGAZİ

AYDIN–2022

ACKNOWLEDGEMENTS

Thank you to my supervisors, Assoc. Prof. Dr. Gözde Başak Öztürk and Asst. Assoc. Prof. Dr. Fatih SOYGAZİ, for your patience, guidance, and support. I have benefited greatly from your wealth of knowledge. I am extremely grateful that you took me on as a student and continued to have faith in me over the years. Thank you to my teammate, Soner CENGİZ. I would like to thank you very much for your great dedication, solidarity, great support, and your encouraging words to achieve our goals in this valuable project we have prepared together.

Thank you to my family, Ayşe, Ali and Cemil Gazi, for your endless support. You have always stood behind me, and this was no exception. Mom, thank you for all of your love and for the support you showed me every day. I would also like to thank my lovely aunt Mahinur Güngör, my uncle Ecevit Güngör, my grandmother Melek Güngör and my cute cousin Beyza Kaya for your big support. I have always felt your warm support by my side. Thank you to my friends Burak Kaan Yılmaztekin, Gülsüm Yaşmak, Hatice Betül Coşar, and Hüsna Şahin, Kim Seok Jin for your friendship, and the great faith and emotional support you have shown me. Your beliefs in me mean a lot to me. And finally, a big thank you to Min Yoon Gi and Kim Nam Joon, and all the other members whose working methods and driven personalities I take as an example, and who I see as role models.

Gamze Gülsün GAZİ

TABLE OF CONTENTS

LIST OF ABBREVATIONS

BIM : Building Information Modelling

FM : Facility Management

GIS : Geographic Information Systems

IoT : Internet of Things

LCA : Life Cycle Assessment

IPD : Integrated Project Delivery

RFID : Radio Frequency Identification

IFC :Industry Foundation Classes

UAV : Unmanned Aerial Vehicles

AR :Augmented Reality

VR :Virtual Reality

LIST OF FIGURES

LIST OF TABLES

ÖZET

BİRLİKTELİK KURAL MADENCİLİĞİ İLE YENİ BİR TREND TAHMİN SİSTEM TASARIMI VE BİLGİ ÇİZGESİ GÖSTERİM: BIM ARAŞTIRMA ALANI UYGULAMASI

GAZİ G. G. Aydın Adnan Menderes Üniversitesi, Fen Bilimleri Enstitüsü, İnşaat Mühendisliği(İngilizce) Programı, Yüksek Lisans Tezi, Aydın, 2022.

Amaç: Çok sayıda bilimsel araştırma ve bilim ve mühendislik alanındaki hızlı gelişmeler, literatürün manuel olarak analiz edilmesini neredeyse imkansız hale getirmektedir. Bu çalışma, gelecekteki çalışmalar için bilgi alanlarını ve trend tahminini analiz etmek için kural tabanlı bir grafik yaklaşımı önermektedir.

Materyal ve Yöntem: Bu çalışmada, Apriori algoritması kullanılarak kural tabanlı bir grafik oluşturulmuştur. Ayrıca aynı anlama gelen ancak literatür için önemli olan ve yazım farklılığı olan anahtar kelimeler bir araya getirilerek tam değerleri elde edilmiştir. Böylece sonuç olarak elde edilen çıktılar daha anlaşılır ve doğru kılınmıştır. BIM ile ilgili veriler, örnek bir araştırma alanı olarak Scopus veri tabanından dışa aktarılmıştır. Çok kapsamlı olduğu için bu araştırma için Scopus veritabanı kullanılmıştır. Analiz arulezViz ile görselleştirilmiş ve kullanıcılar için daha anlaşılır bir sonuç hedeflenmiştir. Makine öğrenme tekniği olan Random Forest ile yakın gelecekteki trendlerin lift değerleri tahmin edilmiştir.

Bulgular: Birden fazla anahtar kelime grubu arasındaki ilişkilendirme kurallarını açıklanmış, ayrıntılı ve nicel analiz çıkarımlarına izin verilmiştir. Aynı anlama gelen ancak yazımları farklı olan bazı kelimeler manuel olarak tek bir anahtar kelimede birleştirilmiştir. Seçilen bazı kurallar makine öğrenmesi tekniğiyle regresyon analizine sokulmuş, ve yakın gelecekteki trendler lift değerleri ile belirlenmiştir.

Sonuç: Bu scientometrik analiz yaklaşımı ile anahtar kelimenin çalışma alanındaki etkinliğinin daha detaylı olarak analiz edilmesi mümkün hale getirilmiştir. Sadece mevcut BIM verisi içerisindeki trendler değil yakın gelecekteki trendlere de ışık tutulmuştur. İleride yapılacak çalışmalarda farklı değişkenler eklenerek analiz zenginleştirilebilir.

Anahtar Kelimeler: Yapı Bilgi Modellemesi, Birliktelik Kuralı Madenciliği, Bilgi Grafiği, Trend Tahmini, Bilgi Yönetimi, Apriori Algoritması

ABSTRACT

A NOVEL TREND PREDICTION SYSTEM DESIGN VIA ASSOCIATION RULE MINING REPRESENTED IN KNOWLEDGE GRAPHS: BIM RESEARCH FIELD APPLICATION

GAZİ G. G. Aydın Adnan Menderes University, Graduate School of Natural and Applied Sciences, Civil Engineering Program, Master Thesis, Aydın, 2022.

Objective: A large number of scientific research and the rapid developments in the field of science and engineering make it almost impossible to analyze the literature manually. This study proposes a rule-based graph approach to analyze the knowledge areas and trend prediction for future studies.

Material and Methods: In this study, a rule-based graph was created by utilizing the Apriori algorithm. In addition, keywords that have the same meaning but are important for the literature with a spelling difference were combined to obtain their exact values. Thus, the outputs obtained as a result are more understandable and accurate. BIM-related data were exported from the Scopus database as an exemplary research area. The Scopus database was used for this research because it is so comprehensive. The analysis was visualized with arulezViz and a more understandable result for users was aimed. Lift values of future trends were predicted with Random Forest, a machine learning technique.

Results: The association rules between multiple word-groups of keywords allowed detailed and quantitative analysis inferences. Some words that mean the same thing but differ in spelling were manually combined into a single keyword. Some selected rules were entered into regression analysis by machine learning technique, and near future trends were determined by lift values.

Conclusion: With this scientometric analysis approach, it is made possible to analyze the effectiveness of the keyword in the field of study in more detail. Not only the trends in the current BIM data, but also the trends in the near future have been shed light on. In future studies, the analysis can be enriched by adding different variables.

Keywords: Building Information Modelling, Association Rule Mining, Knowledge Graph, Trend Prediction, Knowledge Management, Apriori Algorithm.

1.INTRODUCTION

According to international standards, Building Information Model is defined as a common digital representation of the physical and functional characteristics of any built object that provides a reliable basis for decision making (Volk et al., 2014). The growth and complexity of construction projects complicate project management. The interdependencies between different stakeholders such as funding agencies, government agencies, architects, engineers, lawyers, contractors, suppliers, and trades can be cited as an example of complexity (Bryde et al., 2013). Building Information Modelling (BIM) provides a set of interacting policies, processes, and technologies that generate a methodology for managing the life cycle of the project and project data (Succar, 2009). The BIM technology is an accurate virtual model of buildings and is digitally created to be used in architecture, engineering, construction, operations, and facilities management of buildings. The BIM visualization helps architects, engineers, and contractors to identify problems in the phases of design, construction, operation, and facility management (Azhar, 2011). With the effect of developing technology and communication, it is getting harder to manually examine the studies in research areas. For this reason, many various analysis and visualization tools have been developed those researchers can use (Eck and Ludo, 2010; Chen, 2014; Batagelj and Andrej, 2004). In the last decade, technologies such as 3D printing, 3D laser scanning, Artificial Intelligence (AI), point cloud have been integrated into BIM (Siebert and Teizer, 2014; Gheisari and Behzad, 2019; De Melo et al., 2017; Darko et al., 2020). The rapid integration of innovative technologies into BIM has led to a rapid increase in the literature in the field of BIM and its division into more detailed branches. Many scientometric studies have been conducted in order to analyze the research on BIM, which can keep up with technology very quickly, and to identify trends and gaps. This study presents a scientometric analysis of BIM data by creating a knowledge graph and a trend prediction system to analyze the literature in the field of BIM in depth.

In the following sections, a keyword co-occurrence analysis and rule-based keyword analysis of the literature in the field of BIM were conducted. For rule-based analysis, the Apriori algorithm, which is an association rule algorithm, is used. In the methodology section, the stages of the studies are explained. In the findings section, the co-occurrence of keywords has been analyzed and visualized by utilizing traditional scientometric analysis tool. The data is pre-processed and made ready for association rule mining. In addition, at this stage, keywords that are important for the data, have the same meaning but have spelling differences, were combined. This method is aimed to prevent data loss and provide robust analysis. Then, rule-based keyword analysis was performed using the Apriori algorithm. A deeper data analysis is aimed with the Apriori algorithm, which provides association rules by repeating the analysis with the rule based approach. The results of the two analyzes are performed. The discussion section scrutinizes trends and gaps of the BIM research field. The BIM research field literature has been analyzed using two different approaches, one is a traditional computerized literature analyzing method and the other one is the method that was designed within the content of this thesis. the thesis ends with explaining the thesis in summary with the conclusion section. In addition, future studies are highlighted by emphasizing limitations.

1.1.Building Information Modelling

BIM is developing rapidly in the face of the difficulties of project management, of which complexity is increasing due to the growing growth of construction projects, and the use of BIM is becoming more and more widespread (He et al., 2017). One of the main advantages of BIM is that it can contain or link all project-related information to BIM platforms. This great advantage of BIM creates the life cycle of structures and supports IT revaluation in its management (Kalinichuk, 2015). BIM, which keeps up with technology quickly, increased the cooperation between stakeholders by increasing the model dimensions from 3D to nD and played a major role in increasing productivity in the AECO-FM industry (Chatzimichailidou, 2022; Baduge et al., 2022; Pu et al., 2022; Zhai et al., 2022). BIM literature also keeps pace with the developing technology and seeks and develops modern solutions in the face of the difficulties and challenges encountered in the AECO-FM industry (John et al., 2022; Han and Leite, 2022, Bellagarda, 2022). BIM is creating the potential for overcoming many limitations by adopting recent technologies such as the Internet of Things (IoT), Natural Language Processing (NLP), Augmented/Virtual Reality(AR/VR), Lidar, Artificial Intelligence (AI), Blockchain, and GIS for the AECO-FM activities (Zavari et al., 2022; Wang et al., 2022; Zhang et al., 2022; Nassereddine et al., 2022). The mainly focused areas of the BIM research field were explained below to expand the understanding of the subject.

1.1.1.Knowledge Management Via Cognitive Technologies

Increasing complexity due to interdependencies among various stakeholders in construction projects has led to rapid development in the field of information and communication technology [ICT]. An important development in ICT in the construction industry is the proliferation of Building Information Modeling as a new paradigm for computer-aided design (CAD). BIM is a set of policies, processes, and technologies that develop practices for managing design and project data throughout the life cycle of a structure (Bryde et al., 2013). BIM can be used for "visualization," "code reviews," "cost estimating," "construction sequencing," "conflict, interference, and collision detection," and "facility management". The ability of BIM to predict building performance and operation has greatly increased the adoption of BIM. Increased use of BIM enables collaboration between project teams, as well as increased win rates, cost reductions, more efficient time management, and further improved client relationships (Azhar, 2011). The involvement of multidisciplinary teams, the multi-stage project life cycle, and the use of various software and hardware systems/tools have made integration a requirement for the complex structure of the Architecture, Engineering, Construction, Operation, and Facility Management (AECO-FM) industry. Ensuring system integrity is about interoperability (Shen et al., 2010). Although 3D models and applications have recently been used in the AECO-FM industry, collaboration has remained on 2D drawings and documents. The proliferation of computer-aided design (CAD) products, which provide more advanced access to construction processes and increase the level of automation, encourage more collaboration. BIM plays an important role in taking traditional collaboration to a higher level. Apart from 3D drawings, BIM offers capabilities such as providing information between objects and their relationships, reducing technical errors and risks, and facilitating information exchange between stakeholders through the software coded in it (Singh et al., 2011). Building Information Modelling (BIM) provides a set of interacting policies, processes, and technologies that generate a methodology for managing the life cycle of the project and project data (Succar, 2009). The ability of BIM, to visualize projects at an early stage allows designers to easily adjust complex operations such as on-site equipment, procurement, scheduling, and labor. The fact that complex details can be rehearsed at an early stage also minimizes construction risks (Ghaffarianhoseini et al., 2017). Unlike traditional methods, BIM allows project managers to visualize the entire construction process in time and space using nD models. Also as in many other industries, AECO-FM ontology studies have been conducted on various topics such as information management, information management tools, and risk management in the AECO-FM industry (Ding et al., 2016; Tserng et al., 2009; Zhang et al., 2015).Computers can analyze data using ontology and obtain new information. Besides that, many studies have also been conducted such as Industry Foundation Classes (IFC). One of the most important of them is the Industry Foundation Classes (IFC) format. IFC is a digital representation of BIM data that is a primary format for openBIM data exchange and allows building data to be viewed, shared, estimated, simulated, and measured. IFC aims to facilitate interoperability and information exchange between different software programs and stakeholders in the AECO-FM industry through ontology (Lee et al., 2014). In addition, data loss may occur in the acquisition of existing building data since they were not delivered in the BIM format. In this case, laser scanners are used in the AECO-FM industry due to their advantages, such as the ability to scan 3D structures accurately and quickly to capture data, preprocess data, and avoid data loss in BIM modeling compared to traditional modeling (Tang et al., 2010). In addition, GIS systems that capture, store, process, analyze, manage, and present all kinds of geographic data also can be integrated into BIM. Although traditional GIS systems are based on 2D maps, 3D GIS is on the rise nowadays. These systems can store the 3D properties of objects with enhanced functionality (Deng et al., 2016). GIS is used in many areas of the construction industry. Examples of application areas include "transportation", "environmental impact assessment", "urban development", "resource management", "site selection", and "land surveying" (Irizarry et al., 2013; Xu and Volker, 2012). GIS provides an effective basis for planning and monitoring construction activities (Patel et al., 2017). 4D GIS enables the development of the project by visualizing the construction process with its phases, thus facilitating the planning process and the implementation of the plan. The step-by-step visualization of the process reduces the risks and conflicts that may occur during construction (Kumar et al., 2017).

1.1.2.Project Management (PM)

The need for closer collaboration and more effective communication in Integrated Project Delivery (IPD) is driving interest in BIM with new PM frameworks. These features should be documented to provide information to stakeholders and collaborate on the project (Bryde et al., 2013). Recently, specialized planning tools such as energy analysis, structural analysis, scheduling, progress tracking, or site safety have been added to the basic functions (Volk et al., 2014). The building information model, used throughout the project life cycle, characterizes the geometry of building elements, spatial relationships, geographic information, cost calculation and cost control, material information, and time management. As a result, material quantity data and general properties of materials are readily accessible.

All facilities, equipment, and groups of facilities can be represented and managed at a relative scale. Construction documents such as drawings, schedules, and cost control can be easily linked and controlled by stakeholders (Azhar, 2011). BIM is not just a technology. BIM is a process. BIM not only uses three-dimensional models but also enables significant change and management through integrated data throughout the workflow, project delivery, and construction (Hardin and Dave, 2015). BIM enables accurate building sustainability analysis and the creation of high-performance building design by combining multidisciplinary information, building forms, materials, contexts, and comprehensive datasets on mechanical, electrical, plumbing, and other systems into one model. In addition, the coordinated changes and consistency of data save designers from having to create new models and update models. The multidisciplinary structure of BIM enables collaboration in optimizing a low-energy design during the planning and design phase, efficient energy reduction, efficient work during the construction phase, and monitoring and control of building energy consumption (Tang et al., 2017). In practice, Genetic Algorithms (GA) have been used to achieve the goals of BIM and achieve more efficient results. In GA, the process starts with the generation of a solution and evolves using natural evolution steps such as intersection, inversion, and iterative implementation of selection until a suitable solution is found (Bilal et al., 2016).

1.1.3.Lean Construction Management

Lean Construction is a representation of the Toyota Production System (TPS) which has adopted a specific concept and principle in the construction industry. Lean Construction aims to reduce waste, increase value to the customer, and achieve continuous improvement with increasing efficiency. Just as IPD can be BIM-free and fulfills its original purpose with BIM, the implementation of Lean Construction also plays an important role in improving BIM-enabled construction projects (Sacks et al., 2010a). Starting from Deming's 14 approaches based on the quality approach, he has provided a list of principles to analyze the links between BIM and Lean. These are listed under the headings "reduce variability", "reduce cycle times", "reduce lot sizes", "increase flexibility", "choose an appropriate approach to production control", "standardize", "introduce continuous improvement", "use visual management", "align the production system with flow and value", "ensure comprehensive requirements capture", "focus on concept selection", "ensure requirements flow down", "verify and validate" and "convince yourself" (Sacks et al., 2010b). Lean takes a process-based approach to organizing "value streams". At each stage of the process, it is questioned whether the activities add value to the product. If they do not, the activity is declared as waste. This improves the process by reducing the resources spent on wasteful activities (Korb and Sacks, 2018).

1.1.4.Green Approach and Sustainability in Green Buildings

Green buildings can be defined as sustainable buildings or high-performance buildings. Negative impacts of built environments and construction such as noise, dust, water pollution, traffic congestion during construction, consumption of natural and human resources, and environmental impact after construction are well-known factors (Zuo and Zhen-Yu, 2014). Recent studies show that there is an increasing demand for sustainable (also known as "green") buildings that minimize environmental impact. Rising energy costs and increasing global warming are cited as catalysts for demand. Sustainable buildings are also economical, with life cycle savings of about 20% (Azhar et al., 2011). Green BIM is defined as the use of BIM to achieve the performance targets of sustainable buildings. Green BIM is a tool that is used to integrate sustainable structures into the life cycle of these structures in the design industry. Green BIM plays an important role in ensuring energy efficiency (Wong et al., 2015). Designers' focus on more sustainable design led to Life Cycle Assessment (LCA), which captures and evaluates the inputs and outputs of a product throughout its life cycle. The traditional 3D CAD cannot perform sustainable building management. Unlike the traditional approach, BIM represents both the graphical and non-graphical aspects of the building, provides a database, and allows stakeholders to share knowledge before, during, and after construction, and encourages collaboration (Jrade and Jalaei, 2013). Considering these developments, the concept of a digital equivalent of a physical product, "Digital Twin", was introduced at the College of Michigan in 2003. The concept of the digital twin concept model consists of three main parts: 'physical products in Real Space', 'virtual products in Virtual Space', and 'the connections of data and information that tie the virtual and real products together. The amount of information has improved in the virtual space, the products are not only visualized but also behavioral characteristics can be added to test their performance capabilities. This method also provides the ability to create a lightweight version of the virtual product. This allows the required geometries, properties, and attributes to be selected without carrying around unnecessary details. This visibly reduces the size of these models and ensures that the process is maintained more quickly (Grieves, 2015).

1.1.5.Model-based Monitoring and Controlling Via Cognitive Technologies

The ability of BIM, to visualize projects at an early stage allows designers to easily adjust complex operations such as on-site equipment, procurement, scheduling, and labor. The fact that complex details can be rehearsed at an early stage also minimizes construction risks (Ghaffarianhoseini et al., 2017). Unlike traditional methods, BIM allows project managers to visualize the entire construction process in time and space using nD models. nD models capture both temporal and spatial aspects of scheduling, unlike the traditional Gantt chart. This suggests that 4D, which has been used to evaluate and analyze design constructability, sub-contractor coordination, and schedule optimization, can also be used to visualize project risks in time if the process fits well with current risk management (Hartmann et al., 2012). Performing a risk analysis before each activity is important, even if the same activity is repeated continuously. Safety management suffers from the low efficiency of infrequent risk analysis because it requires more effort than people think. In application, Job

Safety Analysis (JSA) is used to predict failures or accidents that may occur due to human or equipment failure and calculate probabilities to effectively allocate safety management resources and plan activities safely. In this way, risk levels can be assessed; it may allow activities to be rescheduled to reduce the peaks at the foreseeable risk levels. In cases where no changes can be made, the necessary safety measures can be planned. Since there is no practical way to track risk waves, construction planners and managers often rely on professional and personal experiences and group assessments (Sacks et al., 2009). Many construction accidents are due to a lack of action, such as worker training, risk source identification and control, safety awareness, and training. However, when examining studies on Virtual and Augmented Reality Construction Safety (VR/AR-CS), it is found that these technologies increase safety and reduce risk on various issues such as worker training, skill teaching, risk identification, and ergonomics (Li et al., 2018). In recent years, the Unmanned Aerial Vehicle (UAV) has been used by researchers (Siebert and Teizer, 2014). UAVs are also used in the construction industry for data acquisition, data screening, and analysis. For direct geographic reference, data from GPS can be used with the timestamps recorded during the flight. Time synchronization of GPS and camera is automatically created. By integrating the data, the point cloud can be transferred to the given coordinate system (Gheisari and Behzad, 2019). UAVs are used in the construction industry in many fields (De Melo et al., 2017; Gheisari and Behzad, 2019). "Mapping of construction sites", "inspection of construction sites" and "surveying of buildings" can be mentioned as UAV applications (Moon et al., 2019; Tuttas et al., 2016; Álvares et al., 2018). Mapping construction sites is a challenging task. Scanning large areas can lead to enormous costs. However, with a drone, the desired area can be easily reached; all information and images can be captured and transmitted in real-time to computer mapping software. Drones are also used for risk and safety management on construction sites. They offer a great advantage by ensuring worker safety and reducing workload. UAV offers great advantages in "building surveys" by providing access to hard-to-control areas. UAV not only saves time and money but is also very reliable in terms of safety on site.

1.2.Scientometric Analysis of BIM

Scientometrics is a branch of informatics that quantitatively analyzes the values in the scientific literature to understand the knowledge structure and emerging trends in a research area. The scientometric analysis uses knowledge mapping methods to reveal relationships, structures, and trends in disciplines by visualizing large amounts of data (Zuo et al., 2021). The results of a scientometric analysis can help those responsible for making decisions about new research areas, potential investment areas, R&D, and operations (Ozcan and Adnan, 2019). As people began to analyze larger networks, researchers and institutions in various disciplines began to use bibliometric network visualization tools, leading to a need for more advanced techniques and tools. For this reason, researchers have developed a variety of tools and techniques that are freely available (Eck and Ludo, 2009). However, there are various limitations that researchers who perform scientometric analysis also face. Yanzhao Li et al. performed cross-checking to avoid data loss or erroneous data for scientometric analysis study (Li et al., 2022). Xiali Xue et al. stated as a limitation of their study is that different content categories may overlap in cluster analysis. The research also stated that the association of keywords is limited to the number of nodes. In addition to these limitations, they emphasized that the lack of a threshold or clipping mode may lead to deviation in the analysis and visualization tool they used in their study (Xue et al., 2021). Tao Wang et al. proposed to reduce the limitations of the analysis tools used in the related study by using innovative technologies such as Artificial Intelligence (Wang et al., 2022). Mirko Locatelli et al. stated that the body of knowledge created in the scientometric analysis is static, which is a limitation (Locatelli et al., 2021). In their study, Kai Sun et al. considered only the basic relational inference instead of a deeper relational inference and explained this as a limitation. For this reason, scholars emphasized that in future studies it is necessary to get to the latent knowledge in the data (Sun et al., 2016). Ozturk (2020a; 2020b; 2021) defined the static nature of the data as a limitation in her study. For this reason, the author stated that a dynamic data import and data visualization could be a solution for current scientometric analysis. The author also stated the 2D visualization is an obstacle to a fast understanding of gap and trend detection

(Ozturk, 2020a; Ozturk, 2021). Many researchers engaged in scientometric analysis and mapping have limited the language of their data to English and they have stated this as a limitation (Wu et al., 2019; Ozturk, 2020b; Baarimah et al., 2021; Wang and Fakhar, 2021; Baffoe and Keith, 2022).

1.3.Scientometric Analysis of BIM Research Field

Many scientometric analyses have been carried out in BIM studies. These analyses aim to determine the progress of studies in the field of BIM, future trends, and gaps in study areas (Wei et al., 2015). In their scientometric analysis study, Li, Xiao, et al. (2017) aimed to determine the BIM knowledge domains, BIM knowledge areas, and the development of BIM knowledge. As a result of this study, a total of sixty fundamental research topics were identified. It is stated that the most important of these fields of study are information systems, 3D/nD modeling applications, design, sustainability, IFCs and interoperability, BIM implementation, multi-dimensional (nD) BIM, real-time communication, and BIM education. In addition to these, 10 knowledge clusters are defined and these clusters are grouped into culture, technology, management, and theory. Jin et al. (2019) aimed to find the main research topics of BIM-based studies in the field of construction engineering and management and to offer research directions for future studies. Jin et al (2019) stated that the current BIM-based studies in the field of construction engineering and management are focused on existing works, and the author envisaged that the focus should be expanded from the pre-construction phases to the construction phases . Vilutiene et al. (2019) aimed to conduct a scientometric analysis on BIM studies in structural engineering. As a result of this study, it is emphasized that BIM could solve the technical problems of structural engineering. Structural components, automation of the assembly sequence, planning and optimization of off-site construction, and dynamic structural health monitoring have been suggested as guidelines for future studies.

Chen et al. (2022) adopted the mixed-review method in this study to review the existing research on the various types of Mega-Infrastructure Projects' (MIP) to get a comprehensive understanding of the current status of the research in the field. The method is defined as the mixed-review analysis and is divided into two quantitative (i.e., bibliometric method) and qualitative (i.e., systematic method) methods. The reason for determining this method is aimed to examine the research in-depth, and it was aimed to eliminate prejudiced and subjective thoughts. In this study, 160 related studies are analyzed, and the research progress of MIPs is critically reviewed. In addition, 3 key findings are indicated as future trends and guided future MIP studies. Marzouk and Elshaboury (2022) conducted an extensive literature review to determine current trends in the embodied energy of buildings worldwide, and to identify future trends and current research gaps. One of the gaps identified in this study is the lack of a complete and standardized database, calculation methodologies, and relevant guidelines for setting evaluation parameters. As a solution to this, controlling the energy effect of building materials at the design stage has been shown. It is also stated as a gap that BIM is not used and combined with other information technologies. It was emphasized that the integration of multiple artificial intelligence technologies should be ensured to provide an accurate and fast energy forecast. In addition, due to the limited analysis of the energy consumption of users and stakeholders, it is recommended to conduct studies in this area to optimize the energy use of residents and stakeholders throughout the life cycle of the building. In addition, various comparative studies and evaluations have been proposed in the fields of "comprehensive rating systems and regulations" and "energy-saving buildings".

1.4. Knowledge Graph

The knowledge graph is a network of a domain, which collects and transmits real-world knowledge with machine learning algorithms. Nodes in the network can be explained as representing the related assets and the edges represent the relationship between these assets (Aidan, 2020). In the Knowledge Graph, applications are built on learning from nodes and connections between nodes, so that Knowledge Graph can benefit a variety of downstream tasks, such as KG-completion and relation extraction. Due to these advantages, it quickly gained great attention (Wang et al., 2017). In practice, knowledge graphs serve as a constantly changing and evolving layer of information within an organization or community. These graphs can be divided into open knowledge graphs and enterprise knowledge graphs. Open knowledge graphs are created by volunteer communities and published online for public benefit. DBpedia, Freebase, Wikidata, and YAGO(Ringler and Paulheim, 2017). Springer, Cham. can be shown as an example of open knowledge graphs. Enterprise knowledge graphs are commercial knowledge graphs used typically within a company. Web search, commerce, and social networks are prominent industries using enterprise knowledge graphs directed edge-labelled graphs are one of the most commonly used graph-structured models for representing data graphs. Also known as a multi-relational graph, this graph is defined as a set of nodes such as {Turkey, Aydin, Istanbul} and a set of directed, labeled edges such as {Istanbul ---City--- > Turkey} between these nodes. In these model knowledge graphs; the nodes are used to represent entities and the edges are used to represent binary relationships between these entities. This model is used to add information to a knowledge graph, typically by adding new nodes and edges, and missing information only requires skipping a particular edge. Another type of model used for knowledge graphs is heterogeneous graphs. Heterogeneous graphs are similar to directed edge-labelled graphs in that each node and edge is assigned a type, but the node type is part of the graph model rather than expressing a specific relationship. Edges are called homogeneous if they are between nodes of the same type, and heterogeneous if they are between different nodes. An advantage of the heterogeneous graph model is that it allows the partitioning of nodes by type for machine learning purposes. The property graph model is introduced to provide additional flexibility when modeling more complex relationships. Property graphs allow a set of property-value pairs and associating a label with both nodes and edges. Property graphs can be converted to/from directed edge-labeled graphs with no information loss. Directed edge-labeled graphs offer a more minimal model, while the property graph model is more flexible. Another knowledge graph model is the graph dataset, which consists of a set of named graphs and a default graph. A named chart is a chart ID and a chart pair, while unidentified charts are referenced by default. One of the uses of graph datasets is to manage and query Linked Data, which consists of interconnected documents of RDF graphs spanning the Web (Aidan, 2020).

1.5.Data Mining

Every day, companies and government agencies store more data in their database systems, but most data of this size cannot be analyzed in a meaningful and efficient way. Data mining is a step in the process of knowledge discovery that automatically extracts hidden, meaningful, and useful patterns in enormous amounts of data (Özçakir and Çamurcu, 2007). Data mining is used in many industries such as banking, insurance, medicine, and retail to reduce data mining costs, improve research and increase sales. Although it was used in the public sector to prevent fraud and waste at first, it has started to be used for performance evaluation and improvement over time. Data mining can be performed on data in quantitative, textual, or multimedia forms, and data mining applications can use various parameters to examine this data. These parameters include attribution, sequence or path analysis, classification, clustering, and forecasting (Pegarkov, 2006). The three steps of data mining are exploration, pattern identification, and deployment. In the first step, the exploration step, the data is cleaned and converted to another format, the important variables and then the nature of the problem-based data are determined. In the second stage, the patterns that make the best prediction are identified and selected to obtain the best prediction. In the last step, the selected patterns are deployed for the desired output. There are many data mining techniques developed recently. These are classification, clustering, association rules, neural networks, regression, artificial intelligence, decision trees, genetic algorithm, and nearest neighbor (Bhardwaj and Saurabh, 2021; Bharati and Ramageri, 2010).

The clustering method is the division of data into groups that are similar to each other and not unlike objects in other groups. The resulting clusters correspond to hidden patterns, the search for clusters is unsupervised learning, and the resulting system expresses a data concept. For this reason, the clustering method is the unsupervised learning of a hidden concept. Three complications arise in the clustering method, namely large data sets, objects with many attributes, and attributes of different types. These difficulties have led to the emergence of powerful clustering methods that can be applied on a large scale (Berkhin, 2006). The most known clustering methods are Hierarchical, Partitioning, Density, Grid-based, and Graph-Based methods (Kameshwaran and Malarvizhi, 2014).

The classification method is the most widely researched data mining technique, which often uses K-Nearest Neighbors (KNN), support vector machines, and Bayes classification algorithms (Taniar, 2008). This technique is particularly suitable for analysis such as fraud and credit risk practices. This method takes place in two stages learning and classification. During the learning phase, the training data is analyzed by the classification algorithm. In the classification phase, the test data is used to predict the accuracy of the classification rules. If accuracy is acceptable, these rules are applied to new data tuples. The classifier-training algorithm uses pre-classified samples to determine the set of parameters required to perform the appropriate classification. The algorithm encodes these parameters in a model called the classifier. Each classifier has a feature that distinguishes it from the others (Bharati and Ramageri, 2010). Types of classification models are classification by decision tree induction, Bayesian classification, K-Nearest Neighbor, neural networks, support vector machines

(SVM), and classification based on associations (Bharati and Ramageri, 2010; Ahmad et al., 2015) These are correctness, time, strength, data size, and extendibility. Correctness classifier is based on the characteristics of how a classifier classifies correctly, and there are some numerical values according to the number of true and false classifications to check the accuracy. The time classifier includes the time that will be spent by the model to build the model and then classify the number of bundles. This determines the computational costs. Strength represents the ability to classify accurately even if the beam has noise. Data size represents the classifiers being independent of the size of the database, making the model scalable. The performance of the created model does not depend on the size of the database. Extendibility means that new features can be added, when necessary, but this feature is difficult to implement (Gorade et al., 2017)

The regression method is a data mining technique used to fit an equation to the data set. Variables whose value is estimated are the dependent variable, and variables whose value is known are the independent variables. In this method, the dependent variable is modeled as a function of several independent variables with corresponding multiple regression coefficients and a constant term (Gupta, 2015). Types of regression methods are linear regression, multivariate linear regression, nonlinear regression, and multivariate nonlinear regression. A linear model can be created using the linear regression method. but in this method, both variables are known. The point here is to trace a line that correlates between both these variables. The linear regression method cannot be used for categorized data, but only for numerical data. However, logistic regression, which takes a non-linear approach, can be used for categorized data. Logistic regression estimates probabilities of occurrence using the logistic regression function (Bharati and Ramageri, 2010).

Neural networks are mathematical model or computational model that exemplifies the functioning of the biological nervous system (Singh et al., 2009). The neural network method is promising in many forecasting and job classification applications due to its ability to learn from data, lack of parametric structures, and generalization (Gaur, 2012). The neural network method mimics the neuron structure of animals. This method is based on the M-P model and Hebb learning rule, its essence is a distributed matrix structure (Ni, 2008). This method can be used to model complex relationships between inputs and outputs or to find patterns in data (Singh et al., 2009). The neural network model can be divided into 3 types as feed-forward networks, feedback networks, and self-organization networks. Feed-forward networks are generally used in forecasting and pattern recognition. In this type, the perception backpropagation model is representative of the network of functions. The feedback network type is often used for associative memory and optimization calculations. Hopfield's discrete model and continuous model are represented in this type. The self-organization networks type is often used for cluster analysis. It accepts this type of adaptive resonance theory (ART) and the Kohen model as representative (Gaur, 2012). The advantages of neural networks can be listed as noise tolerance, high accuracy, independence from previous assumptions, ease of maintenance, applicability to parallel hardware, and the ability to continue without any problems in case of failure of any neural network element due to its parallel nature. However, there is no general method to determine the optimum number of neurons in this method and it is difficult to choose a training set that fully describes the problem (Singh et al., 2009).

Artificial Intelligence (AI) attempts to apply similar processing to human thought (Sharma, 2014). AI uses external data such as large to perform given tasks at peak performance. The advantages of AI methods can be listed as follows: Speed in decision-making, problem-solving ability, disregard of emotions in decision-making processes, easy dissemination of information, ensuring the persistence of data and prevent data loss, following a similar path as the thought processes of humans, completing tasks faster than humans, working with minimal errors, working simultaneously with different the ability to work and complete complex tasks easily, finding unexplored areas and being programmed for a long time. However, AI cannot guarantee the optimal solution and cannot be used in general cases that cannot be represented in the dataset. In addition, some of its decisions cannot be explained logically and may cause wrong solutions in small malfunctions. In addition, it requires a lot of money and time (PK, 1984).

1.5.1.Association Rule Mining

Association rules analyze data from the past, identify the co-occurrence of patterns in that data and show the possibilities of co-occurrence in the future. It is an approach that supports future studies by examining data from the past. The association rule can be expressed as X → Y. If X is present in a transaction, then Y is highly likely to be present in the transaction (Hong and Yeong-Chyi, 2008). Let's assume that there are minimum support, minimum confidence, and minimum lift values set by the user. In the first iteration phase of the algorithm, each item is determined as a 1-itemset candidate. Among the detected items, only candidates with values equal to or greater than minimum support, minimum confidence, and minimum lift are brought forward for the next iteration, and those left behind are pruned. Then the candidates meet the minimum values form the 2-itemset set. In the next iteration, 2-itemset elements that meet the minimum values are detected. In this iteration, 3-itemset will be created using the merge and prune step. In this iteration, each item is combined with itself, producing 3-itemset candidate members from 2-itemsets. Items that do not meet the minimum values are then pruned. 3-itemset is created with the remaining members. The next step would be to join the 3-itemset with itself and prune any items that do not meet the minimum values. 4-itemset is created with candidates who have met the minimum values. The algorithm stops working when it reaches the most common set of items. The purpose of the association rule is to find common item sets in large datasets (Bharati and Ramageri, 2010). Association rules detect unknown relationships and produce results that can form the basis for decision making and prediction (Al-Maolegi and Bassam, 2014). Types of association rules are multilevel association rules, multidimensional association rules, and quantitative association rules (Bharati and Ramageri, 2010). The Apriori algorithm, one of the best-known algorithms for generating association rules, is an effective algorithm for learning association rules between items and finding frequently used items in knowledge graphs. As items appear more frequently, they are expanded into larger clusters. Items designated as favorites are then used to define association rules (Jia et al., 2018). This works in two main steps. In the first step, frequently used item sets are created. In this step, various combinations of items are formed and item sets that are not equal to or greater than the specified minimum support are pruned to find itemsets of dimensions-1 (1-itemsets). Item sets of dimension-2 (2-itemsets) are created from the common 1-itemsets created in the second step, and item sets that do not meet the minimum support value are pruned. This process continues until no item set meets the minimum support value (Castro et al., 2018). The Apriori algorithm is simple and one of the most classical association rules known (Singh et al., 2012). The Apriori algorithm reduces the candidate itemsets below the support limit determined at each stage, providing a satisfactory performance gain (Chee et al., 2019). However, when the total number of item sets increases, it is necessary to generate a large number of candidate items, then rescan the entire database and validate a large number of candidate articles. For this reason, this algorithm is not useful for exceptionally large data sets (Wu et al., 2008).

2.METHODOLOGY

Many articles, seminars, and workshops have claimed that BIM is a catalyst for change that can reduce industry fragmentation, increase efficiency/effectiveness, and reduce the prohibitive cost of poor interoperability. These claims involve a variety of thought constructs derived from organizational studies, information systems, and regulatory fields, abbreviated as much as possible. These differences and scope highlight the absence and need for a research framework for domain knowledge organization, which requires systematic exploration of the BIM research field (Succar, 2009). This study presents a rule based knowledge graph development with association rule mining for BIM research in the AECO-FM industry. In this study, the Scopus database is used to obtain the data about the research area of BIM. The reason for using Scopus is that it provides the most comprehensive data for BIM literature (Ozturk, 2020b), as it allows to examine the writing habits and keyword choices of all authors in the BIM field. A browser bot code was written in python for faster and more complete data exportation. In this way, the bot logged into the Scopus database with the identity information of the researcher. Data with desired limitations were searched by using the search input in the Scopus advanced search section. Journals and subjects that are not relevant to the topic were removed from the search results. The bibliometric Search Input for the BIM research field is given in Figure 1. Thus, cleaner, and more homogeneous data was obtained. The data obtained as a result of the search were filtered and grouped on a yearly basis. The first 2000 articles of the article groups filtered by years were exported as CSV files. The reason for choosing the first 2000 articles is that the Scopus database allows a maximum of 2000 articles in each export selection. For this reason, in this study, the data were grouped according to years, divided into small pieces, and exported in pieces. Thus, the written code was able to export almost all of the data without losing any. Data downloaded in batches based on year were combined into a single CSV file. Keyword co-occurrence analysis, document-based citation analysis, and author-based citation analysis are obtained by utilizing the traditional scientometric analysis tool. The results are visualized, and keyword values are tabulated manually, which is another limitation of the traditional method. The Traditional Scientometric Analysis methodology flow chart is given in Figure 2.

{BIM} OR {building information modeling} OR {building information modelling} AND NOT {bearingless induction motor} AND (LIMIT-TO (LANGUAGE,"English")) AND (LIMIT-TO (SRCTYPE,"j")) AND (LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (SUBJAREA,"ENGI") OR LIMIT-TO (SUBJAREA,"COMP") OR LIMIT-TO (SUBJAREA,"SOCI") OR LIMIT-TO (SUBJAREA,"ENVI") OR LIMIT-TO (SUBJAREA,"BUSI") OR LIMIT-TO (SUBJAREA,"ENER") OR LIMIT-TO (SUBJAREA,"MULT") OR LIMIT-TO (SUBJAREA,"ARTS") OR LIMIT-TO (SUBJAREA,"DECI") OR LIMIT-TO (SUBJAREA,"ECON") OR EXCLUDE (SUBJAREA,"PHYS") OR EXCLUDE (SUBJAREA,"CENG") OR EXCLUDE (SUBJAREA,"BIOC") OR EXCLUDE (SUBJAREA,"MEDI") OR EXCLUDE (SUBJAREA,"CHEM") OR EXCLUDE (SUBJAREA,"AGRI") OR EXCLUDE (SUBJAREA,"PSYC") OR EXCLUDE (SUBJAREA,"PHAR") OR EXCLUDE (SUBJAREA,"HEAL") OR EXCLUDE (SUBJAREA,"NEUR") OR EXCLUDE (SUBJAREA,"IMMU") OR EXCLUDE (SUBJAREA,"NURS") OR EXCLUDE (SUBJAREA,"DENT") OR EXCLUDE (SUBJAREA,"VETE")) AND (EXCLUDE (EXACTSRCTITLE,"Plos One") OR EXCLUDE (EXACTSRCTITLE,"International Journal Of Molecular Sciences") OR EXCLUDE (EXACTSRCTITLE,"Neuropsychological Rehabilitation") OR EXCLUDE (EXACTSRCTITLE,"International Journal Of Building Pathology And Adaptation") OR EXCLUDE (EXACTSRCTITLE,"Environmental Toxicology And Pharmacology") OR EXCLUDE (EXACTSRCTITLE,"Fish And Shellfish Immunology") OR EXCLUDE (EXACTSRCTITLE,"Archives Of Toxicology") OR EXCLUDE (EXACTSRCTITLE,"Bioconjugate Chemistry") OR EXCLUDE (EXACTSRCTITLE,"Biomaterials") OR EXCLUDE (EXACTSRCTITLE,"Bioresource Technology") OR EXCLUDE (EXACTSRCTITLE,"Artificial Cells Nanomedicine And Biotechnology") OR EXCLUDE (EXACTSRCTITLE,"BMC Systems Biology") OR EXCLUDE (EXACTSRCTITLE,"Journal Of Computational And Theoretical Nanoscience")

OR EXCLUDE (EXACTSRCTITLE,"Nature") OR EXCLUDE (EXACTSRCTITLE,"Environmental Toxicology") OR EXCLUDE (EXACTSRCTITLE,"Environmental Science And Pollution Research") OR EXCLUDE (EXACTSRCTITLE,"Human And Experimental Toxicology") OR EXCLUDE (EXACTSRCTITLE,"Marine Policy") OR EXCLUDE (EXACTSRCTITLE,"IEEE Transactions On Antennas And Propagation") OR EXCLUDE (EXACTSRCTITLE,"IEEE Transactions On Energy Conversion") OR EXCLUDE (EXACTSRCTITLE,"Management Learning") OR EXCLUDE (EXACTSRCTITLE,"Ocean Engineering") OR EXCLUDE (EXACTSRCTITLE,"Plos Computational Biology") OR EXCLUDE (EXACTSRCTITLE,"Advances In Mechanical Engineering") OR EXCLUDE (EXACTSRCTITLE,"Atmospheric Environment") OR EXCLUDE (EXACTSRCTITLE,"Chemical Engineering Journal") OR EXCLUDE (EXACTSRCTITLE,"Computational Materials Science") OR EXCLUDE (EXACTSRCTITLE,"Journal Of Materials Chemistry A") OR EXCLUDE (EXACTSRCTITLE,"Man In India") OR EXCLUDE (EXACTSRCTITLE,"Arabian Journal Of Geosciences") OR EXCLUDE (EXACTSRCTITLE,"Electric Power Components And Systems") OR EXCLUDE (EXACTSRCTITLE,"Engineering Analysis With Boundary Elements") OR EXCLUDE (EXACTSRCTITLE,"Scientific Reports") OR EXCLUDE (EXACTSRCTITLE,"Proceedings Of The National Academy Of Sciences Of The United States Of America") OR EXCLUDE (EXACTSRCTITLE,"Journal Of Geophysical Research Space Physics") OR EXCLUDE (EXACTSRCTITLE,"IEEE Access") OR EXCLUDE (EXACTSRCTITLE,"Agro Food Industry Hi Tech") OR EXCLUDE (EXACTSRCTITLE,"International Journal Of Biological Macromolecules") OR EXCLUDE (EXACTSRCTITLE,"Chemosphere") OR EXCLUDE (EXACTSRCTITLE,"IEEE Transactions On Industrial Electronics") OR EXCLUDE (EXACTSRCTITLE,"IEEE Transactions On Magnetics") OR EXCLUDE (EXACTSRCTITLE,"Science") OR EXCLUDE (EXACTSRCTITLE,"Journal Of Commonwealth Literature") OR EXCLUDE (EXACTSRCTITLE,"Journal Of Alloys And Compounds") OR EXCLUDE (EXACTSRCTITLE,"Journal Of Physical Chemistry C") OR EXCLUDE (EXACTSRCTITLE,"International Journal Of Environmental And Science Education") OR EXCLUDE (EXACTSRCTITLE,"Proceedings Of SPIE The International Society For Optical

Engineering") OR EXCLUDE (EXACTSRCTITLE,"Energy Conversion And Management") OR EXCLUDE (EXACTSRCTITLE,"Engineering Failure Analysis") OR EXCLUDE (EXACTSRCTITLE,"Journal Of Chemical Information And Modeling") OR EXCLUDE (EXACTSRCTITLE,"Heliyon") OR EXCLUDE (EXACTSRCTITLE,"ACS Nano"))

Figure 1. Bibliometric Search Input

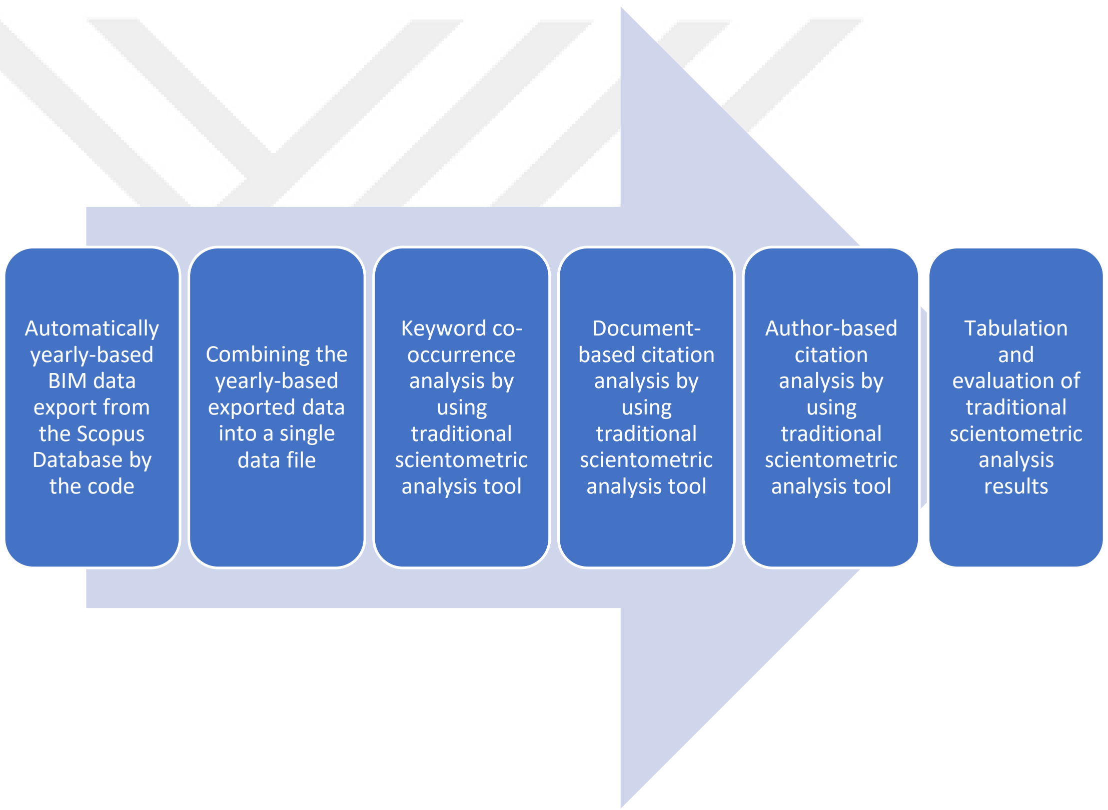


Figure 2. Traditional Scientometric Analysis Methodology Flow Chart

NLTK, NumPy, pandas, Matplotlib, Apriori, and arulesViz libraries were used in this research. First of all, these libraries are downloaded (Bird et al., 2009; Harris et al., 2020; Jeff Reback et al.,2020; John, 2007; Toivonen, 2017; Hahsler, 2021). The NumPy library is the basic package for performing scientific calculations in Python. NumPy allows to operate large sized mathematical data. Pandas is a library written in Python that provides data analysis and associated manipulation. It allows for the analysis of data in various formats, tabulated and separated by commas. Matplotlib is a cross-platform of NumPy as a data visualization and graph plotting library. Matplotlib provides rich graphics and visualizations for numerical data for users. The data is normalized and converted to a format that the computer can understand using the Natural Language Processing Tool Kit (NLPTK) on Python. Keywords with different spellings but the same meaning that are important in BIM are equated with a single keyword, which increases the accuracy of the analysis. The Apriori algorithm in the Apyori library, an association rule mining approach, is used to convert the normalized data into a rule based knowledge graph and make it understandable to the user. the total link strength values of the traditional scientometric analysis tool and the support values obtained in the Apriori algorithm were correlated to prove the accuracy of the chosen algorithm. The data obtained from the traditional scientometric analysis tool and Support and Total Link Strength values obtained by utilizing the Apriori algorithm are correlated using Equation 6 and Equation 7 (Eck and Ludo, 2009). The traditional tool used in this study stands for "visualization of similarities". Equation 6 expresses the direct similarity measure function. *O* is an occurrence matrix of order m × n where columns of **O** correspond with the objects of which to analyze the co-occurrences. *O* is equal to 1 if there is an object at the intersection of the *k* row and *i* column and zero otherwise. Let's assume $c_{ij}$is the element in the *i*th row and jth column of *C*. For $i \neq j$, *cij* equals the number of co-occurrences of objects *i* and *j*. If $i = j$, the value of *cii* is found.

$$c_{ij} = c_{ii} \sum_{k=1}^{m} O_{ki} O_{kj}$$

(2.1.)

S_i defines the total number of cooccurrence of *i*. Therefore S_i can be described as

$$s_i = c_{ii} \sum_{k=1}^{m} O_{ki}$$

(2.2.)

Secondly, it can be expressed as

$$s_i = \sum_{j=1, j \neq i}^{n} c_{ij}$$

(2.3.)

The domain of $S(c_{ij}\ , s_i, s_i)$ equals DS = {($c_{ij}\ , s_i, s_i$) ∈ R^3|0 ≤ c_{ij} ≤ min(s_i, s_j) and s_i, s_j> 0}, where R denotes the set of all real numbers. The range of $S(c_{ij}\ , s_i, s_i)$ is a subset of R and $S(c_{ij}\ , s_i, s_i)$ is symmetric in s_i ands_j, that is, $S(c_{ij}\ , s_i, s_i) = S(c_{ij}\ , s_i, s_i)$ for all ($c_{ij}\ , s_i, s_i$) ∈ D_S. Equation 6 is for correlation in this study

$$S_A(c_{ij}, s_{i,} s_j) \frac{c_{ij}}{s_i s_j}$$

(2.4.)

and Equation 7 is for correlation

$$S_A(c_{ij}, s_{i,}s_j)\,\frac{c_{ij}}{\sqrt{s_i s_j}}$$

(2.5.)

Correlation results are obtained as 0.815542338694835 for Equation 6 and 0.9513988018458107 for Equation 7. Thus, it is decided that the Apriori algorithm is the right algorithm for rule-based analysis.

2.1. Definitions

1.Support: The support (supp) value usually represents the importance of an association pattern in the transaction population. The probability that the association pattern exists in the process population gives support (Tan et al., 2004).

supp(X) = P(X)

2.Confidence: The confidence (conf) value of the *X Y* rule is the probabilities in the transaction population which contain both *X* and *Y* (Taniar, 2008).

conf(X → Y) = supp(X →Y)/supp(X) = P(X∪Y)/P(X) = P(Y / X)

3. Lift: The lift value calculates the number of times the relationship model *X* and *Y* occurred more than expected. This factor is used in data mining to measure deviation from statistical independence (Tan et al., 2004).

lift(X → Y) = lift(Y → X) = conf(X →Y)/supp(Y) = conf(Y → X)/supp(X) = P(X ∪Y)/ (P(X)P(Y))

If the lift value is greater than 1, it means that the two items are more likely to be seen together. If the lift value is less than 1, it indicates that the two items are more likely to be seen separately from each other. If the lift value is equal to 1, it means that there is no relationship between the two items. The knowledge graph with the desired minimum values for support, confidence, and lift values is visualized using arulesViz, which is an R extension package that applies various visualization techniques to explore association rules. The values of the visualized rules are tabulated. Some of the association rules of some trends determined in BIM data were taken as an example and the lift values of these rules over the years were recalculated with the Apriori algorithm. Trend prediction of BIM data is executed by using random forest regression approach of machine learning over these values. Random forest regression is a supervised learning algorithm that combines predictions from multiple machine learning algorithms to make a more accurate prediction from a single model(Ho, T. K., 1995).

By comparing the results of the two analysis approaches, key areas of investigation in the field of BIM are discussed and research areas for future studies are suggested. The rule based knowledge graph via association rule mining for knowledge management and trend prediction system design methodology flow chart is given in Figure 3.

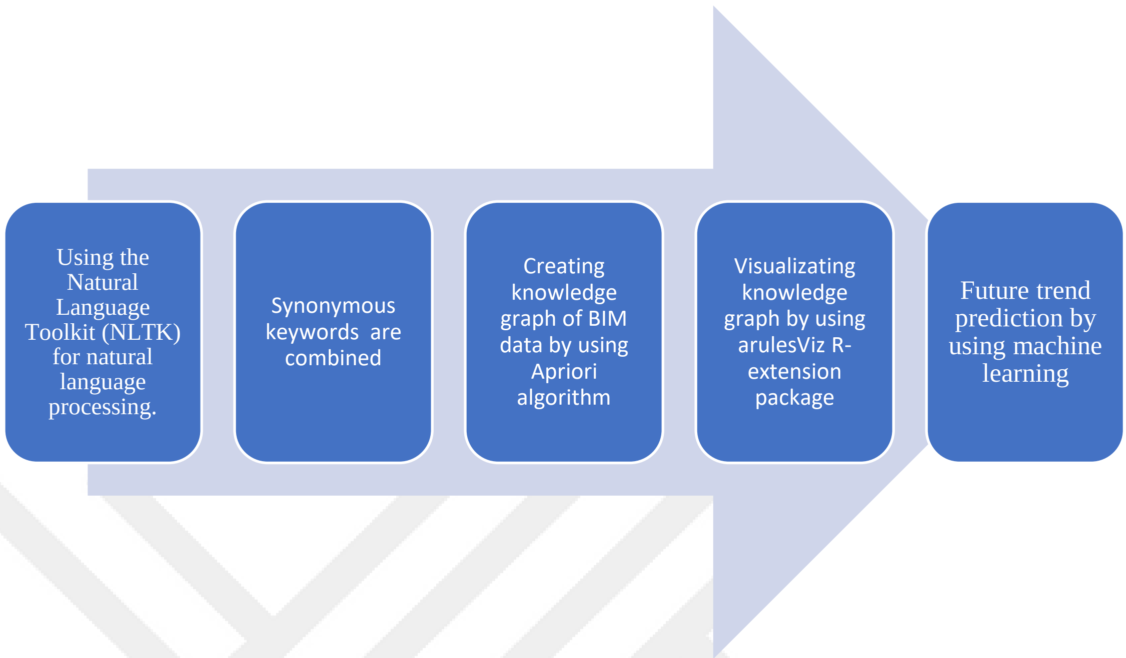


Figure 3. The Rule Based Knowledge Graph Via Association Rule Mining for Knowledge Management and Trend Prediction System Design Methodology Flow Chart

3.FINDINGS

3.1.Exporting Data Automatically from Database

In this study, a Python code has been written to retrieve data from Scopus faster and more accurately. The Scopus database allows exporting only the first 2000 of the resulting articles in Advanced Search at a time. For this reason, if the articles to be analyzed are more than 2000, all articles must be exported manually as 2000 in number. Due to this situation, a python code has been written that automatically classifies and exports the data. The written code automatically enters the Scopus user information and enters the system, thus providing access to the advanced search. The bot automatically searches the results by entering the predetermined search input into the search section. To create a search input first, "BIM" and "building information modeling" and keywords are written into braces ({}) and by using Boolean Operators (AND, OR) [{BIM} OR {building information modeling} OR {building information modelling}] on Advanced Section on "Article title, Abstract, Keywords" section, a researched made. Results are limited in Subject areas to "Computer Science", "Engineering", "Social Sciences "and "Energy", "Business Management and Accounting", "Multidisciplinary", "Environmental Science", "Arts and Humanities", "Decision Sciences", and "Economics, Econometrics and Finance". Document type is limited to "Article". Language is limited to English. The reason the language is limited to English is that English is the language of science (Ozturk, 2020a). The journals which are not related to the AECO-FM industry are excluded. Finally, more detailed restrictions were added to find only construction-related BIM papers. The results obtained as a result of the search are filtered by code on a yearly basis. Articles classified by year are exported as groups. Since the number of articles in most article groups did not exceed 2000, articles could be exported. Article data files classified by year are combined into a single CSV file. The final number of articles that were used for scientometric analysis and mapping was 12185.

3.2.Scientometric Analysis and Mapping of BIM Research Field

3.2.1. Scientometric Keyword Co-Occurrence Analysis and Mapping by Utilizing Traditional Scientometric Analysis Tool

In this analysis, the co-occurrence analysis of keywords of the BIM research area was conducted for the period between 1960 and 2022. Co-occurrence analysis of keywords in the research area BIM was performed for the period between 1960 and 2022. Both author and index keywords are selected for the analysis. The purpose of including index keywords in the analysis is to prevent the possible loss of keywords on the subject, whether used or neglected by the author. The "Full counting" option was used for the keywords in the analysis. The "Full counting" option considers the number of a keyword in all documents. Thus, the count of a keyword in a document indicates that the document is targeting that keyword. 51415 keywords were analyzed. The min occurrence number is set to 120. 130 keywords met the threshold. 77 keywords were selected for the analysis.

The result of the co-occurrence keyword analysis is mapped and visualized in a network diagram in Figure 4. In this diagram, each node represents a keyword. The size of the nodes varies according to the value of the occurrence of the keyword. The lines between the nodes define the relationship between the keywords. The value of this relationship is defined by the link value. The link value indicates the strength of the link between the two keywords. The distance between keywords in the visualization indicates the co-occurrence relationship of the keywords. The closer two keywords are to each other, the more likely they are to be found together. This means that the relationship between them is stronger. The colors of the nodes represent the cluster to which the nodes belong. In this analysis, the keywords are divided into five clusters. The result of the analysis is a list of keywords, given cluster names in Table 1 such as "sustainability-based", "semantics-based", "adoption-based", and "management-based”. The total Link Strength represents the strength of the links between a keyword with other keywords. The occurrence value indicates how often the keyword occurs in articles. The higher the occurrence value, the more actively the keyword is used in the research area. The average citation value shows the average number of citations of the articles in which a keyword is mentioned, and the average normalized citation value shows the effectiveness of the articles in which the keyword is mentioned in the research area. The average publication year indicates how much new a keyword is in the research field. The network visualization map of the keyword co-occurrence analysis in the BIM research field is given in Figure 4. The Keyword Co-occurrence Analysis results are given in Table 1. The Building Information Modelling focused network visualization map of the scientometric keyword co-occurrence analysis in the BIM research field is given in Figure 5.

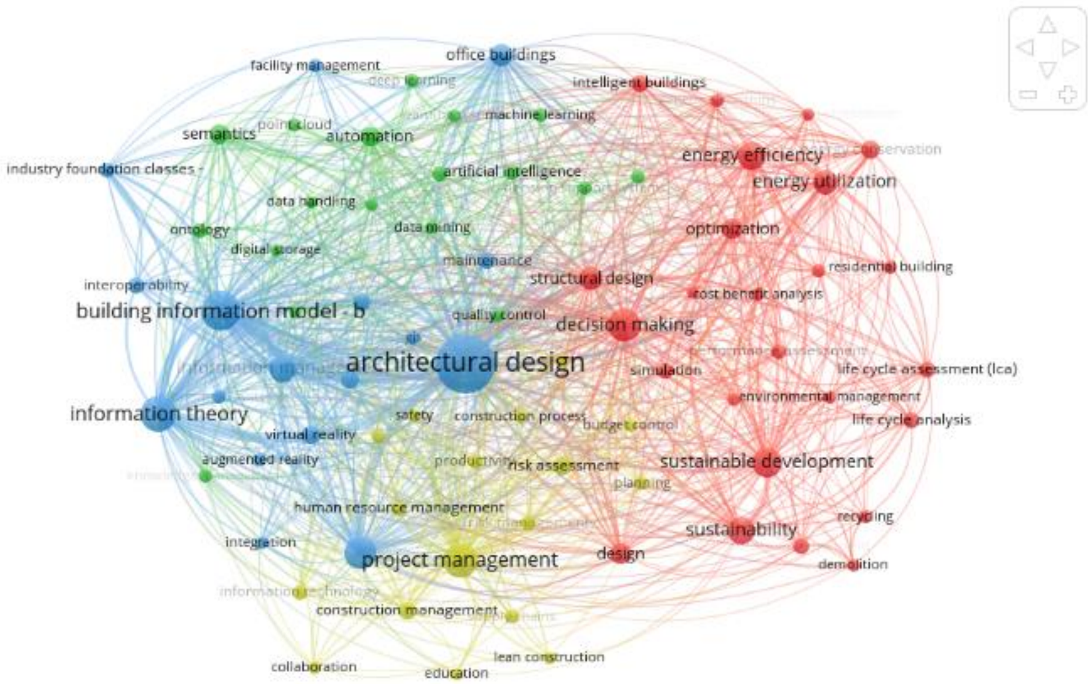


Figure 4. Network Visualization Map of the Keyword Co-occurrence Analysis in the BIM Research Field

Table 1. Keyword Co-occurrence Analysis in BIM for AECO-FM industry

Keyword in the BIM Research Field	Link	Total Link Strength	Occurrence	Average Citation	Average Normalized Citation	Average Year Published
Cluster 1 – Sustainability -Based						
Benchmarking	68	342	126	21.50	1,44	2018,32
Cost Benefit Analysis	67	549	170	15.88	1,23	2018,52
Decision Making	76	2732	881	16.63	1,30	2018,68
Demolition	52	368	121	23.37	1,77	2019,26
Design	72	1009	343	21.36	0,98	2016,80
Energy Conservation	64	856	279	13.86	1,06	2018,39
Energy Efficiency	71	1946	651	15.21	1,19	2018,59
Energy Utilization	73	1665	488	14.75	1,22	2018,79
Environmental Management	66	511	136	19.26	1,47	2018,50
Genetic Algorithms	62	539	160	18.26	1,37	2018,32
Intelligent Buildings	70	986	241	21.36	1,43	2018,39
Life Cycle Analysis	57	614	214	13.09	1,17	2019,15
Life Cycle Assessment (LCA)	54	684	201	20.26	1,78	2018,99
Multiobjective Optimization	59	504	121	19.47	1,62	2018,90
Optimization	73	1027	376	19.03	1,25	2018,16
Performance Assessment	68	552	190	17.08	1,35	2018,70
Recycling	55	410	143	24.38	2,07	2018,97
Residential Building	56	464	154	11.51	1,12	2018,86

Sensitivity Analysis	62	428	146	15.68	1,34	2018,99
Simulation	72	543	224	17.02	0,97	2017,93
Structural Design	76	1741	472	22.48	1,17	2017,23
Sustainability	73	1202	581	11.90	1,12	2019,11
Sustainable Development	74	2054	719	17.44	1,39	2018,81
Uncertainty Analysis	64	376	126	14.94	1,25	2018,76
Waste Management	66	560	202	21.30	1,60	2018,80
Cluster 2 – Semantics-based						
Artificial Intelligence	70	609	193	16.40	1,42	2018,49
Automation	75	979	323	20.94	1,39	2018,16
Data Acquisition	65	376	123	21.37	1,40	2018,41
Data Handling	72	436	128	19.16	1,23	2018,02
Data Mining	70	456	158	14.03	1,28	2018,99
Decision Support Systems	72	657	169	19.41	1,40	2018,13
Deep Learning	59	321	149	9.25	1,39	2020,36
Digital Storage	65	407	133	19.74	1,41	2018,35
Forecasting	64	412	158	11.18	1,37	2019,47
Internet Of Things	63	421	177	20.03	2,50	2019,63
Knowledge Based Systems	63	417	122	19.80	1,15	2017,21
Knowledge Management	62	430	160	19.17	1,12	2017,78
Learning Systems	66	385	130	18.82	1,71	2018,98
Machine Learning	68	439	170	10.48	1,49	2020,08
Ontology	65	591	191	25.21	1,20	2017,33
Point Cloud	51	284	147	25.91	1,35	2018,24
Quality Control	69	400	153	17.70	1,21	2018,60
Semantics	72	1082	313	20.81	1,29	2017,90
Cluster 3 - Adoption-based						
Architectural Design	76	9191	2763	24.20	1,34	2017,73

Augmented Reality	56	375	170	24.21	1,42	2018,7
Building Information Model – BIM	76	4772	1245	26.43	1,48	2017,53
Building Information Modelling	76	2742	849	19.57	1,18	2018,06
Computer Aided Design	71	770	261	23.66	0,96	2015,98
Facility Management	67	599	153	25.09	1,43	2017,54
Gis	64	304	166	18.41	1,07	2017,94
Industry Foundation Classed – Ifc	65	861	198	30.82	1,61	2017,26
Information Management	76	2145	572	20.95	1,37	2017,75
Information Systems	67	517	148	22.90	1,34	2016,70
Information Theory	76	4435	1033	28.41	1,49	2017,47
Integration	61	429	139	20.26	1,15	2016,96
Interoperability	69	765	218	31.68	1,42	2016,54
Maintenance	3	744	231	11.22	0,97	2018,78
Office Buildings	75	160	389	21.67	1,38	2018,21
Virtual Reality	71	608	287	19.72	1,35	2018,15
Visualization	71	811	273	27.52	1,22	2017,34
Cluster 4 – Management-based						
Accident prevention	62	410	147	28.10	1,78	2018,01
Budget control	70	396	126	15.12	0,98	2018,18
Collaboration	46	297	131	33.51	1,38	2016,35
Construction management	73	738	250	22.27	1,08	2017,54
Construction process	74	395	137	22.92	1,16	2017,63
Education	53	253	121	16.47	0,77	2016,78
Human resource management	69	619	195	20.24	1,30	2018,23
Information technology	64	538	177	36.93	1,18	2014,76

Lean construction	55	267	131	22.53	1,18	2017,93
Planning	71	467	152	19.05	1,22	2017,94
Productivity	67	436	178	17.37	1,11	2017,90
Project management	76	2730	988	20.44	1,21	2017,84
Risk assessment	73	741	301	14.39	1,23	2018,91
Risk management	65	485	178	12.15	1,03	2018,58
Safety	59	316	127	19.92	1,25	2018,02
Scheduling	66	387	129	19.91	1,17	2017,6
Supply chains	60	379	164	19.60	1,61	2018,38

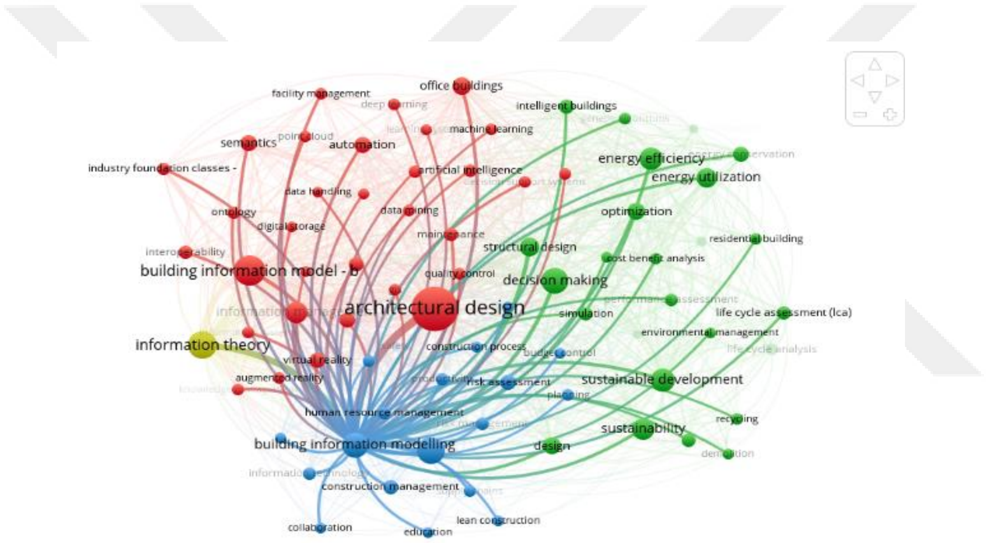


Figure 5. Building Information Modelling Focused Network Visualization Map of **the** Scientometric Keyword Co-occurrence Analysis in the BIM Research Field

The keywords for node-based scientometric analysis “architectural design”, “building information model”, “information theory”, “project management”, “decision making”, and “building information modeling” have the highest occurrence and total link strength values. Also, these keywords have the highest and same link value. In this analysis, “architectural design” has the highest occurrence and total link strength values. After the “architectural design”, the keyword “building information model” has the highest value of occurrence and total link strength. This shows that how comprehensive the approach of “building information model” is in architectural design. The research in the field of BIM focuses more on the adoption-based areas. As evidenced by the link values prove, it can be observed that studies in BIM focus on project management and information theory-based decision making. “Information theory”, “collaboration”, “interoperability”, and “industry foundation classes – ifc” keywords are the keywords with the highest average citation. Information theory-based interoperability studies in the field of BIM aim to improve collaboration between disciplines and stakeholders through the increased implementation of BIM. Knowledge obtained with BIM through industry foundation classes (IFC) is shared, managed, and stored throughout the building's lifecycle from design to demolition. “Internet of things,” “recycling”, “Life Cycle Assessment – LCA”, and “accident prevention” are determined as the keywords with the highest normalized citation value. This is because the biggest and main purpose of BIM studies in the AECO-FM industry is to take the most accurate measures and decisions against all the risks and problems that may be encountered during the life cycle phases of the building. “Deep learning” “machine learning”, “internet of things”, “forecasting” and “digital storage” are frequently used keywords in BIM studies recently. By integrating BIM and the Internet of Things, objects, the environment, the building itself, BIM and personnel will be connected. With radio frequency identification and wireless sensor technology on the Internet of Things, a continuous flow of information is provided for machine learning and deep learning mechanisms in BIM. The Internet of Things can provide the up-to-date data needed for the prediction mechanism of BIM. Thus, the risks, problems, and situations that the building may encounter during its life cycle can be predicted more accurately and quickly. Although the BIM literature initially focused on three-dimensional designs, as technology developed, and systems developed were adapted to BIM tools. Thus, Sustainability-based studies in the literature have begun to focus on the more effective use of information accumulated in Building Information Models. Using the data BIM collects throughout the life cycle of buildings enables more sustainable, economical, and safer design, operation, and maintenance. The increase in complex structures and the difficulty of collecting information in various areas such as energy efficiency, cost control, and safety triggered for need various optimization and decision-making mechanisms for BIM tools. The processing of information collected in building information models by decision support system tools has been of significant help in forecasting in building life cycle. Therefore, studies in the near future tend to the fields of optimization, intelligent buildings, and energy conservation according to node-based scientometric analysis. Semantic-based studies in the field of BIM have focused on automation, artificial intelligence, decision support systems, and ontology to make sense of and utilize the data collected. By accelerating the integration of multi-purpose optimization tools in BIM, stakeholders can be presented with different alternatives by using the information collected in the building information models. However, more detailed research is needed to improve the performance of capturing, understanding, processing, and decision-making in a single environment for all the information collected in different building information models during the life of the building. Therefore, there is a need for advanced integration of BIM and data mining and deep learning techniques. For this reason, studies in the fields of data handling, data mining, deep learning, digital storage, and forecasting are increasing. However, there is a need for studies on adapting systems such as the Internet of Things (IoT) and Point Cloud to meet data needs through the regular data flow. As it is known, the adoption-based studies in the field of BIM are mostly focused on architectural design, building information model, information management, and information theory. Building information models, which offer multidimensional solutions in architectural design and information management, have been rapidly adopted due to the ease of information management. However, the increase in different methods and tools thanks to the rapidly developing technology has increased the need for studies in the fields of interoperability and IFC. In addition to the recent increase in studies on these issues, adoption-based BIM studies will also show a trend in these issues in the near future. In addition, with the widespread use of BIM in the management of life cycles of buildings, besides architectural design, studies in the field of Facility Management have gained momentum recently.

3.2.2. Document-based citation analysis and mapping by Utilizing Traditional Scientometric Analysis Tool

Mapping results from the BIM research for the AECO-FM industry document-based citation analysis are shown in Figure 6. Each circle represents a document. The size of the circles is directly proportional to the contribution of the document they represent to the research area, i.e., the larger the circle, the larger the contribution of the document to the research area. The distance between the circles and the links between the circles show the relationship between the documents and the mutual citations. The link, citation, normalized citation, and publication year values obtained as a result of the analysis are given in Table 2. The link value indicates the number of links an article has with other articles. The citation value indicates the number of citations a document receives, and the normalized citation value indicates the normalized value of the number of citations a document receives. In this analysis, the minimum citation of documents is limited to 35. Unrelated documents are eliminated and a total of 12 documents are analyzed. The three top articles with the highest citation value are Azhar (2011), Succar (2009), and Bryde (2013). The documents with the highest normalized citation values are Siebert et al., (2014), Azhar (2011), and Becerik-Gerber et al., (2012).

Azhar (2011) aimed to provide useful information to the AEC industry practitioners who want to adopt BIM technology into their projects. Azhar (2011) provides a detailed discussion of the applications, trends, benefits, risks, and future challenges of BIM. Under the title of BIM applications, the usage purposes and key benefits of BIM are briefly explained. BIM's purpose of usage is explained in many groups as "visualization", "fabrication drawings", "code reviews", "cost estimating", "construction sequencing, conflict, interface, and collision detection", "forensic analysis", and "facility management" separately. The key benefit of BIM is represented as the correct geometric model of the building in an integrated data environment. In addition, more design possibilities, provision and management of life cycle data, better production quality, and better customer service are specified as key benefits of BIM. The role of BIM in the AEC industry and its current and future trends are discussed based on the results of a 2008 survey on the use of BIM in the AEC industry and a 2007 survey on the value and success-contributing factors of using BIM. Future challenges are discussed by considering shortcomings in the handling and use of BIM. The results of these reports show that BIM users are aware of the benefits of BIM and that the AEC industry wants to consider BIM as a whole. As a result of the survey findings, Azhar (2011) stated that BIM users want BIM not only as a CAD tool with powerful documentation and visualization capabilities but also as a BIM application that supports multiple design and management processes. Four different case studies were examined for the benefits of BIM in the project planning, design, pre-production, and construction phases, and the benefits of BIM were discussed. BIM risks are divided into legal and technical risks, and the key risks in each category are briefly discussed. This study presented the benefits, risks, and future challenges of BIM, with concrete cases, reports and survey results useful for the AEC industry players looking to incorporate BIM into their projects. The report and survey results discussed have also clearly revealed the expectations of BIM users and the AEC industry.

Succar (2009) introduces the BIM framework to allow systematic exploration of different areas of BIM, to define information components, and draw expanding boundaries. It also provides an ontology and visual information models to represent domain concepts and their relationships. Succar (2009) explained the BIM framework under three main headings as BIM fields, BIM maturity levels, and BIM lenses. Succar divides the fields of activity of BIM into 3 main parts "The BIM Technology Field", "The BIM Process Field", and "The BIM Policy Field". The class defined as Technology Field includes the equipment, software, hardware, and networking systems necessary for the productivity, efficiency, and profitability of the AEC industry. In addition, organizations that produce software and produce various software solutions are also included in these groups. Process Field includes all players who supply, design, construct, manufacture, use, maintain and manage buildings throughout their life cycle. Process field cluster includes players which do not produce construction products, but responsible of preparing practitioners, delivering research, distributing benefits, allocating risks, and minimizing conflicts within the AECO industry. In the maturity stages framework, transformational or radical changes separate the stages. Also, the steps are incremental.

Maturity stages contain components introduced in the technology, process, and policy fields. The beginning of the BIM stages is Pre-BIM, and the long-term goal is Integrated Project Delivery (IPD). Apart from this beginning and conclusion, the stages are divided into three object-based modeling, model-based collaboration, and network-based integration. In the lens framework, lenses are divided into A "Macroscopic Lens", "Mesoscopic Lens", and "A Microscopic Lens". Also, the differences between lenses and filters are discussed. A BIM ontology is presented to reduce the complexity of the BIM framework, facilitate knowledge acquisition, and validate the themes of the BIM framework. In addition, information visualization with different model formats has been made to represent fields, concepts, relations, and ontological infrastructure in the most accurate way. With this study, Succar presented frameworks that will facilitate the work of future BIM studies and presented the BIM framework to researchers more systematically by creating and visualizing an ontology. This work has also become an important reference for future BIM framework studies (Sanhudo et al., 2020; Amirebrahimi et al., 2016; Boje et al., 2020; Cerovsek, 2011).

Bryde (2013) analyzed cases published in academic journals and via the internet to determine the benefits of BIM and its positive and negative effects on construction projects. Secondary data is used in this article. The reason for this is that there are benefits such as self-reporting of secondary data, thus reducing information bias. In this study, a list of success criteria is created to measure the success of both the project and the project management. This ensured that not only the success of the project but also that of the project management could be analyzed quantitatively. These criteria are created according to the Project Management Institute (PMI) Project Management Body of Knowledge (PMBOK) Knowledge Areas (PMI, 2008)(Webster, 1993). In this way, the presentation of the analysis and its results are made more comprehensive and understandable. All the case studies were considered in the success criteria, and these case studies were examined in detail by adding a plus 1 point for each positive benefit and -1 point for each negative benefit. As a result of this study, it was determined that BIM provides the most benefit in Cost criteria, followed by Time, Communication, Coordination Improvement, and Quality. It has also been stated that the negative effects of BIM are generally caused by training and software issues. It was emphasized that the benefit obtained from BIM will be increased by adapting BIM to the entire life cycle of structures, and therefore the damage caused by training and software problems will be minimized. It has been explained that once people are trained and computer equipment is upgraded, the negative aspects of BIM will not be left. This study demonstrates that BIM is compatible with PM literature, with integration, better collaboration, and construction project team coordination. This study demonstrated the potential benefits of BIM through case report analysis and charted a realistic course for future work in the field of BIM.

BIM has been adopted in the construction industry in visualization, analysis, control, and similar areas. Becerik-Gerber (2012) discuss how useful BIM will be in facility management (FM) and the synergy between FM applications and BIM. This study used personal interviews and online survey results as data on the grounds that there was very little empirical data. This survey and interviews, it is aimed to determine the industry's level of interest and experience regarding the use of BIM in FM. Different understandings were encountered as a result of interviews with experts from different backgrounds working in various companies in the FM industry to design the survey questions. As a result of the interviews, many functions are performed manually in the FM industry, and it was concluded that many changes would be made, and FM efficiency would be increased by adapting BIM to the FM industry. The purpose of the online survey study is to determine the extent to which BIM is implemented in FM, identify FM organizations' plans to adopt BIM, and find areas where BIM can add value to FM organizations. The results of the survey concluded that more than half of BIM users in the FM industry are aware of the benefits of BIM in various FM organizations and plan to use BIM in the future. In addition, multiple BIM areas in FM organizations have been selected as areas that can potentially add value. Among these, there are "locating building components", "facilitating real-time data access", and "visualization and marketing" as the most selected areas. In addition, survey results show that BIM users are more optimistic than non-BIM users in BIM areas that will add potential value. Concerns about the use of BIM in the FM industry were also identified in the comments received from the participants, as well as the survey answers. Most reviews focused on the barriers to implementing BIM in FM. According to the data obtained from the survey results, the current and future areas of FM applications in which BIM is effective and in which areas they can be useful are discussed in detail. Thus, BIM has shown a clear direction for future work in the FM industry. In addition to these, some technological and organizational difficulties in BIM application are listed. He introduced the continuity of the life cycle with a nongeometric data structure proposed by the article and discussed the roles and responsibilities of stakeholders for data provision. Thus, Becerik-Gerber (2012) highlights the benefits of BIM in FM, the responsibilities of stakeholders, the challenges to be addressed in the future, and the gaps that need to be filled, as it presents future studies.

BIM studies in the AECO-FM industry include extensive literature dating back to the 1960s. BIM has led to the emergence of many fields in the literature, thanks to its design and facilitation of construction operations and its easy adaptation technology (Bryde et al., 2013; Succar, 2009; Azhar, 2011; Siebert and Teizer, 2014; Becerik-Gerber et al 2012; Barlish and Sullivan, 2012; Zhang et al., 2013; Eadie et al., 2013; Schlueter and Thesseling, 2009; Xiong et al., 2013; Thormark, 2002; Gu, N and London, 2010). In document analysis, the most cited studies in the BIM literature consist of articles that introduce expanding BIM scope, explore BIM's adaptation, and provide the benefits that BIM can bring to the AECO-FM industry. In these studies, the benefits of BIM adaptation were discussed, supported by case studies and surveys, and the AECO-FM industry encouraged its stakeholders to BIM. In addition, these studies shed light on the identified challenges and future directions for future BIM research. The articles examined in this analysis cover various topics such as the benefits of BIM, BIM frameworks, BIM adaptations, the relationship of BIM to life cycle, energy efficiency, safety, laser scanning technologies. Among the studies in the field of BIM, the most cited documents have generally focused on the definition, understanding, benefits, and limitations of BIM. Other studies following these studies have examined the various uses of BIM throughout the building life cycle. The studies in the field of BIM are very detailed and broad. These articles have become a starting point for studies in the field of BIM, which is very easy to keep up with technology, with the education and technology-oriented topics it proposes for future research. However, there is a need for interoperability studies in the field of BIM. More detailed studies are needed to improve performance in capturing, understanding, processing, and decision-making in a single environment for all information collected in different building information models during the building life cycle.

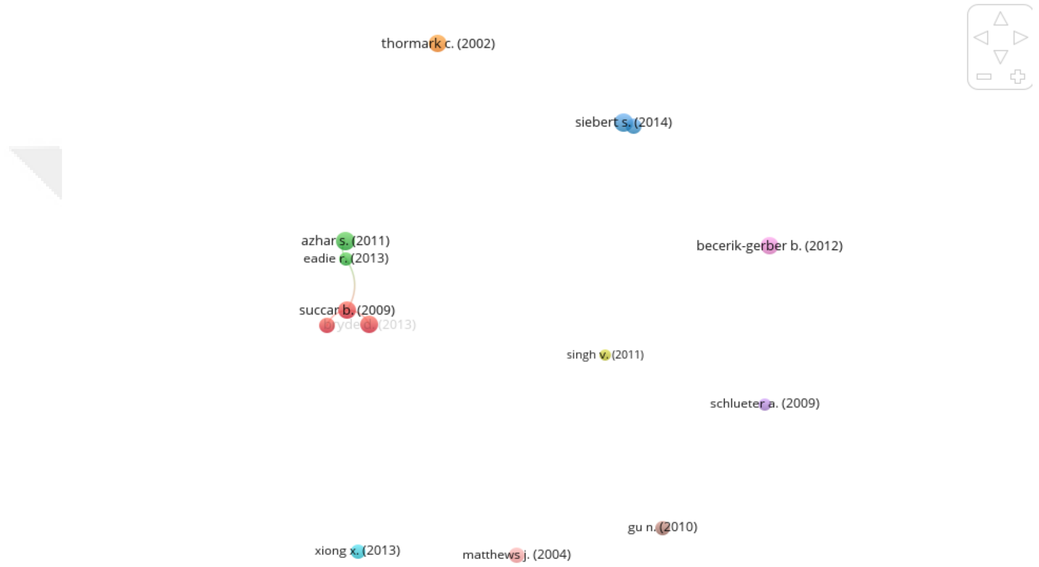


Figure 6. Network Visualization Map of Document-based Citation Analysis of BIM Research Field

Table 2. Document-based Citation Analysis in BIM for AECO-FM industry

Documents	Link	Citation	Normalized Citation	Average Publication Year
Barlish K. et al., (2012)	1	377	11,31	2012
Bryde D. et al., (2013)	1	620	14,11	2013

Succar B., (2009)	1	764	14,23	2009
Siebert S. et al, (2014)	1	487	15,84	2014
Zhang S.et al, (2013)	1	471	10,72	2013
Azhar S., (2011)	1	938	15,84	2011
Eadie R. et al., (2013)	2	402	0,76	1995
Singh V. et al., (2011)	0	367	6,20	2011
Schlueter A. et al., (2009)	0	394	7,34	2009
Xiong X., (2013)	0	409	9,30	2013
Thormark C., (2002)	0	471	14,07	2002
Gu. N., (2010)	0	476	8,97	2010
Becerik-Gerber B. et al., (2012)		478	14,35	2012

3.2.3.Author-Based Citation Analysis and Mapping by Utilizing Traditional Scientometric Analysis Tool

Mapping results of author-based citation analysis of the BIM research field are shown in Table 3 and in Figure 7. Each circle represents an author. The size of the circles is directly proportional to the contribution of the author they represent to the research area, i.e., the larger the circle, the larger the contribution of the author to the research area. The distance between the circles and the links between the circles show the relationship between the author and the mutual citations. For the author-based citation analysis, the minimum number of documents is set to 5 and the minimum number of citations is set to 1500. The result of this analysis is that 12 authors have exceeded these limits.

The Average Normalized Citation value indicates the average normalized number of citations received by the author. “Wang X.”, “Wang J.”, and “Li H.” have the most normalized citation value among the most influential authors in this area of research, as shown in Figure 7. One of Wang X's most highly cited articles is his study with Wang J and Wu P. Wang X. aimed to review the concept of three-dimensional printing, its features, and its applications in the construction industry, and identify 3D challenges in the construction industry. Top citation articles discussed integrating the Geographic Information System (GIS) and Augmented Reality/Virtual Reality(AR/VR) into BIM. In these studies, he benefited from case studies and aimed to integrate these technologies into the field of BIM by creating frameworks and ontologies. Wang X. discussed the benefits of adopting technologies such as GIS, and AR/VR into BIM fields, the difficulties encountered in the integration process, and the working aspects of the studies in the BIM. Wang J. has studied the integration of BIM into buildings, Infrastructure Industries, and FM organizations in his most cited articles. In addition, he discussed the adaptation of technologies such as Lidar, GIS, and AR in different areas of BIM and the contribution of these technologies to the purposes of BIM. Li H. has researched the implementation of BIM in construction projects by supporting it with case studies in his studies, where he received the highest citation. Various effects of BIM in the construction industry are discussed in these studies.

The FWCI value is the ratio between the number of citations received by an article or group of articles and the expected number of citations, depending on the type of article and the average number of citations per article in the year of publication. The expected number of citations is determined by the expected rate of actual citations for a particular document, documents of similar age, topic, and publication type. An FWCI value above 1 indicates that more citations that the document received than expected. Although the FWCI value has greater variation than normal for newer articles, it will normalize over time as the FWCI value is updated. The FWCI value indicates the author's prestige and citation performance in the research field. Cheng J. C. P., Wang J., Wang X., Tezier J., Zhang S., and Li H. are who received the highest FWCI value in their studies in the field of BIM, within the topic "Construction Sites; Information Modeling; Point Cloud" (Scopus-Defined) in terms of citation performance compared to other authors. In addition, the author who has the most studies on this topic is Li H. He has carried out various studies in the field of safety and construction, such as sensor-based safety management, vision-based detection and visualization applications, and digital printing. Areas such as integration challenges, safety problems in the construction of large and complex structures, and time and cost savings triggered growth in the adaptation of various technologies, ensuring construction safety by using different technologies, and information management. Deep Learning, Prefabricated Building, Internet of Things, Industrial Robot, Occupational Health and Safety, Terrestrial Laser Scanning, 3D Laser Scanning, Activity Recognition, and Convolutional Neural Network are determined as areas with very high growth values in "Construction Sites; Information Modeling; Point Cloud" topic. Adaptations of emerging technologies in the face of the challenges and issues faced in the construction industry in the field of BIM are seen among the emerging research areas. In this topic, Wang X. Wang J. conducted computer vision techniques digital twins, RFID, and machine learning applications in the construction industry. In addition to these studies, he has worked on BIM implementation on different subjects. Cheng J.C.P. has conducted various studies in the field of computer vision, deep learning techniques, and the adaptation of various technologies to the AEC industry and construction safety. Teizer J. has conducted cloud, virtual reality, and various technology-based construction safety studies in this topic area. In addition, he conducted Internet of Things (IoT) studies in the fields of data integration and lean construction education. Zhang S. conducted various studies in the field of automated systems for safety in construction. He also carried out some studies on BIM integrations with technologies such as semantic web and RFID.

Li H., Love P. E. D., and Wang X. are the authors with the highest total number of articles. Wang X, Li H., and Wang J. are the authors with the highest total number of articles in the BIM field. The H-index measures the authors' influence on the literature. The h-index is based on both the number of published articles by the author and the number of citations these articles have received. While the total number of articles shows the author's contribution to the whole literature, the number of articles related to the BIM field shows the author's tendency to the BIM research field. Love P. E. D., Li H. Teizer J., Sacks R. are the authors with the highest H-index value. The Link value indicates the given author's relationship with other authors. Total link strength indicates the total strength of an author's relationship with all other authors. Wang X., Wang J., Li H., and Sacks R. are the authors with the highest total link strength. Wang X., Wang J., and Li H. have the highest average normalized citation value. FWCI, Number of Articles, Number of Articles Related to Research Area, H-Index, Link, Total Link Strength, and Normalized Citation values of author analysis were presented in Table 3.

Table 3. Author-based Citation Analysis in the BIM Research Field

Research Theme	Author	FWCI	Number of Articles	Number of Art. Related to Research Area	H-Index	Link	Total Link Strength	Average Norm. Citation
BIM Integrations (Red-coded)	Azhar S.	1.22	141	10	34	1	77	42.97
	Cheng J. C. P.	1.59	87	51	34	1	161	107.33
	Lu W.	1.33	193	73	39	1	158	137.81
	Wang J.	1.59	77	121	24	1	288	247.13

	Wang X.	1.59	348	160	53	1	373	329.24
BIM Implementation (Green-Coded)	Eastman C. M.	1.50	158	25	35	1	177	62.66
	Sacks R.	1.41	147	50	41	1	214	87.47
	Tezier J.	1.59	172	26	43	1	181	78.54
	Zhang S.	1.59	17	51	12	1	176	74.10
BIM Practices (Blu-coded)	Becerik-Gerber B.	1.23	149	22	39	1	78	57.08
	Li H.	1.59	534	135	64	1	239	157.91
	Love P.E.D.	0,69	466	66	75	11	116	115.96

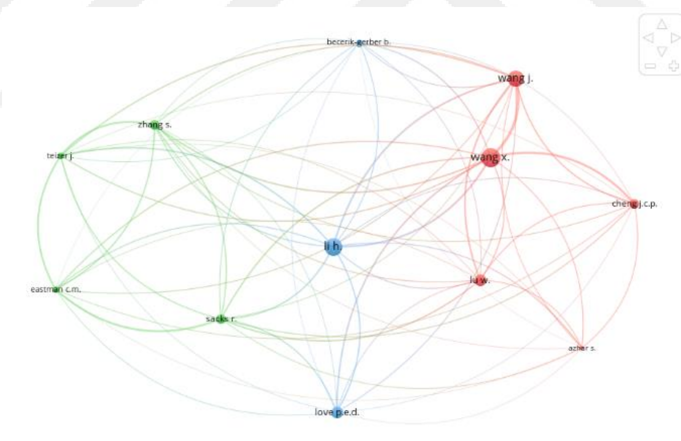


Figure 7. Network Visualization Map of Author-based Citation Analysis in BIM Research Field

In the author analysis, authors are grouped under the main titles "BIM Integration", "BIM implementation”, and "BIM practices". The most influential authors in BIM research are Love P.E.D., Li H., Teizer J., Sacks R, Lu W., and Becerik-Gerber B. In addition, BIM integrations and BIM practices were ranked as more noteworthy by researchers. With the development of BIM studies as technology evolves, the integration of BIM with new technologies has become more important than the practical application of BIM. For this reason, the tendency of studies in BIM is mainly toward the relationship between new technologies, BIM, and the adaptation of new technologies to BIM studies in real life.

In this study, keyword-based, document-based, and author-based citation analyses were performed by utilizing a traditional scientometric analysis tool, which exhibits a node-based similarity approach in these analyses. In the analysis, only the links between two nodes representing keywords, documents, and authors could be examined, and inferences were drawn from the values formed by these links. Moreover, the values of words that have the same meaning, but different spellings are distributed across multiple nodes, so the actual values of some keywords could not be determined. For this reason, the keywords with the highest value in this feature were selected. This limitation led to the loss in values and accuracy deviations in the analysis. For this reason, a rule-based prediction system has been developed in order to find solutions to these problems and for obtaining robust analysis results in the BIM research field.

3.3.Data Preprocessing

Data preprocessing has an important place in data mining. At this stage, the data is transformed into more useful condition. Data may contain irrelevant or meaningless parts. These unnecessary parts of the data that may slow down or hinder the data analysis (unnecessary words, spaces, columns, etc.) were cleaned. With data transformation, the values in the data are converted into appropriate forms according to the applied mining process. Data reduction methods are used in big data preprocessing due to concerns such as storage efficiency, data storage, and analysis costs. This method is aimed to reduce the data capacity appropriately. However, there was no need to use the data reduction method in this study.

Columns containing keyword information of studies in BIM CSV file were checked and repeated columns were removed. The keyword list is consolidated. The Natural Language Toolkit (NLTK) is used for natural language processing. "stopwords", "tokenize", "sent_tokenize", "word tokenize", modules of the NLTK library are imported. With ”stopwords”, words that do not have meaning and therefore want to be ignored are filtered out. Tokenizing allows working with relatively more coherent and meaningful pieces of text by dividing the text into words or sentences. In this study, word tokenizing is applied. The Numpy library, the Pandas library, pyplot, and arulesViz libraries are imported.

Index and Author keywords in the retrieved data are in different columns. To analyze all keywords, “Index Keywords” column and “Author Keywords” column are combined and named “All Keywords”. All Keywords are converted into a sentence by using the “labels_to_sentences” function. The text of the data is tokenized into words and word groups by every semicolon “;” in the text. All letters are converted into lower case. White spaces, accent marks, and other diacritics were removed. Numbers are converted to words, the number of punctuations, all letters were converted to lower case, abbreviations. A function named “clear_kwlist” was defined and keywords expressing spaces were removed from the created keyword list with this function. The keywords in the "Keywords_" column in the data set were clarified with the relevant functions and defined as a list in a variable named "kws".

3.3.1. Combining Synonymous Keywords

In keyword-based scientometric analysis, words with the same meaning have different spellings may result in data loss in the analysis. In addition, this situation may cause the decrease in the accuracy rate of the results. For this reason, in this study, certain synonymous keywords that are valuable for the literature in the field of BIM are combined. In the keywords in the "kws" list, there are words that have the same meaning but are different in spelling. For this reason, the support value of a keyword is divided (scattered) into keywords in different spellings. The words that have the same meaning with different spellings, which are important in the BIM literature discussed in this study, are equated to a single keyword, and the division of the support values they contain is prevented by the "map_list" function. The abbreviation BIM is not equated to any word as it can mean both building information modelling and building information model. For this reason, the BIM abbreviation is not synchronized to any keywords.

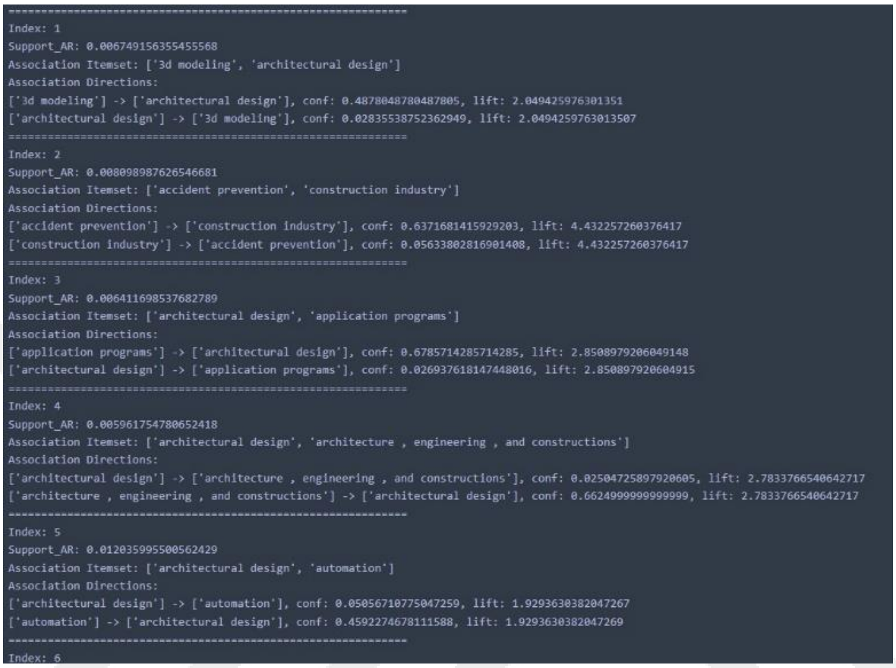
```
keyword_map = {
    "building information modelling": "building information modelling (BIM)",
    "building information modeling": "building information modelling (BIM)",
    "building information modeling (bim)": "building information modelling (BIM)",
    "building information modelling (bim)": "building information modelling (BIM)",
…
```

3.3.2.Excluding Keywords Irrelevant to Data

In the visualization package ArulesViz, the user can extract the desired words from the visualization. However, the eliminated words are not removed from the association rules, so the user can see the relationships of these words on the scatter plot. A first step is required to select unwanted words and remove them from the analysis. For this reason, a function has been added to the preliminary part of the data that allows the user to extract the keywords that cause pollution of the data. The function named "filter_keywords_from_txt" takes the kwlists and the location of the .txt file with the unwanted keywords as inputs. The path of the .txt file with unwanted keywords can preferably be left empty. If the user leaves it empty, the keywords in the kwlists will not be filtered. Keywords in the .txt file with the extension specified in the function are assigned to the "unwanteds" list. The superfluous whitespaces of the keywords in the "unwanted" list, which should be excluded by the user, are cleaned up with the "preprocess_word" function. Then the lists with all keywords determined as kwlists and the “unwanted” list with unwanted words are passed to the "remove_unwanted_list" function as inputs. In the "remove_unwanted_list" function, the keywords in kwlists are compared with the keywords in the “unwanted” list. Matching words are excluded from the kwlists lists.

3.4.Creating A Knowledge Graph by Utilizing Apriori Algorithm

To apply the Apriori algorithm to BIM data, the Apriori function of the "Apyori" library on GitHub is used. Apriori function gives a list as input and frozenset as output. The frozen output includes that the association rules in the data are formed between the keywords and the support, confidence, and lift values of the created rules. But the output is not useful and understandable by researches. Two functions are defined as “frozenset2list” and “printAssociation” to make the output as frozenset more meaningful. With "frozenset2list", the output is made into a list. With the "printAssociation" function, the output of the Apriori function is printed in a meaningful way. To use the Apriori function, the variable “kws”, minimum support value, minimum confidence value, and minimum lift value are given. By giving these values as low as possible, it is aimed to examine as many keyword relationships as possible in the literature. After that, with the "printAssociation" function, keyword relations and values are output as in Figure 8.



```
==============================================================
Index: 1
Support_AR: 0.006749156355455568
Association Itemset: ['3d modeling', 'architectural design']
Association Directions:
['3d modeling'] -> ['architectural design'], conf: 0.4878048780487805, lift: 2.049425976301351
['architectural design'] -> ['3d modeling'], conf: 0.02835538752362949, lift: 2.0494259763013507
==============================================================
Index: 2
Support_AR: 0.008098987626546681
Association Itemset: ['accident prevention', 'construction industry']
Association Directions:
['accident prevention'] -> ['construction industry'], conf: 0.6371681415929203, lift: 4.432257260376417
['construction industry'] -> ['accident prevention'], conf: 0.05633802816901408, lift: 4.432257260376417
==============================================================
Index: 3
Support_AR: 0.006411698537682789
Association Itemset: ['architectural design', 'application programs']
Association Directions:
['application programs'] -> ['architectural design'], conf: 0.6785714285714285, lift: 2.8508979206049148
['architectural design'] -> ['application programs'], conf: 0.026937618147448016, lift: 2.850897920604915
==============================================================
Index: 4
Support_AR: 0.005961754780652418
Association Itemset: ['architectural design', 'architecture , engineering , and constructions']
Association Directions:
['architectural design'] -> ['architecture , engineering , and constructions'], conf: 0.02504725897920605, lift: 2.7833766540642717
['architecture , engineering , and constructions'] -> ['architectural design'], conf: 0.6624999999999999, lift: 2.7833766540642717
--------------------------------------------------------------
Index: 5
Support_AR: 0.012035995500562429
Association Itemset: ['architectural design', 'automation']
Association Directions:
['architectural design'] -> ['automation'], conf: 0.05056710775047259, lift: 1.9293630382047267
['automation'] -> ['architectural design'], conf: 0.4592274678111588, lift: 1.9293630382047269
--------------------------------------------------------------
Index: 6
```

Figure 8. Rule-Based Graph Output for BIM Research Field Application

3.5.Visualizing Knowledge Graph - Arulesviz

The R extension package arulesViz was used to visualize the knowledge graph. Support, lift, and confidence values are assigned to the keywords. The goal is to comprehensively analyze the literature in the field of BIM by keeping these values as small as possible. Thus, not only the most effective keywords but also keywords that require more work and will determine future trends are included in the analysis visualization. The minimum confidence value was set to 0.002, the minimum support value was set to 0.0011, and the minimum lift value was set to 0,0015. The rule based knowledge graph of BIM data visualization with arulesViz was shown in Figure 9.

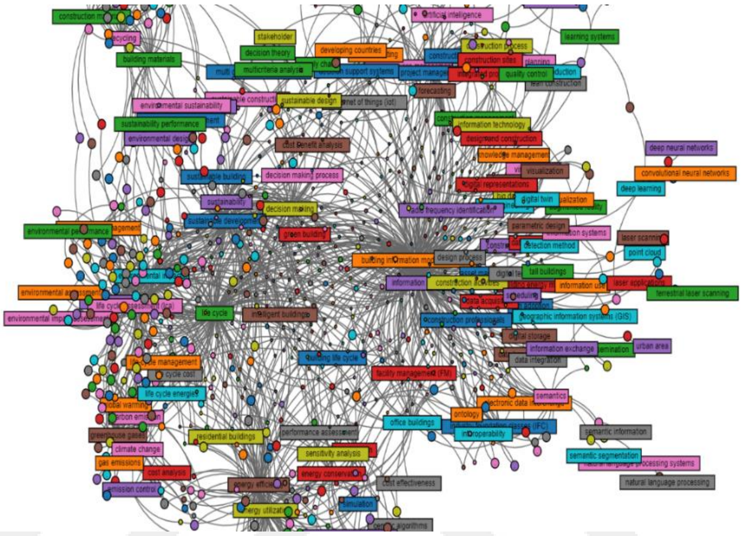


Figure 9. Rule-Based Graph Visualization of BIM Data

The desired information can be viewed by clicking on the keywords. By double-clicking on the keywords, the detailed values of the Confidence, Support, and Lift values of the keywords and their relationship with other keywords can be accessed by Scatter plot. By double-clicking on the desired keyword, the scatter plot of that keyword can be observed. Scatter plot visualization is especially useful for visualizing and analyzing big data (Hahsler, 2021). The rules can be examined more easily by zooming in on the desired points. The darkness of the dots indicates the strength of the lift value. The darker the dot, the higher the lift value of the relationship it is expressing. The lines between the keywords show the relationship between the keyword and other keywords. With this visualization library, the visualization can restrict the visualization by adding the desired constraints to the values. In addition, by selecting the desired keywords, only the relationship between the desired keywords can be visualized and examined. To analyze the knowledge graph visualized in big data more easily, the width, charge, and link distance values can be changed by changing the dimensions of the graphic and accessing the graphic information can be facilitated. The lift value indicates how much or less than expected the association rule it represents is encountered in the data. This value shows the subjects that attract more attention and are studied more than expected in the field of the study analyzed in scientometric analysis.

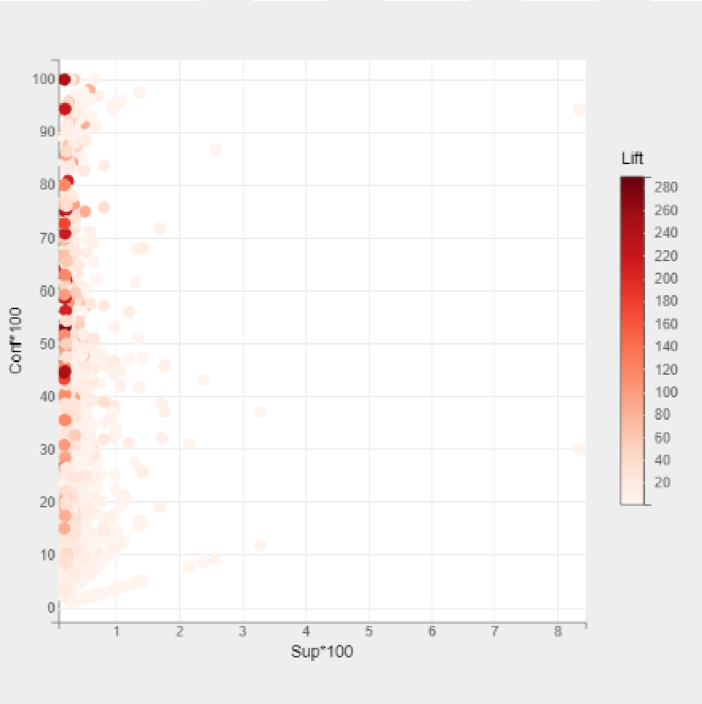


Figure 10: Scatter Plot of the Keyword of "Building Information Modelling"

The rules with the highest lift value of the keywords identified as latest trends in the node-based analysis were listed in Table 4 and visualized in Figure 11 below.

Table 4. The Rules with the Highest Lift Value of the Keywords Identified as Recent Trends in the Node-Based Analysis

Association rules	*Confidence*	*Support*	*Lift*
Optimization			
Optimization algorithms → optimization	0,481	0,001	14,249
Optimization modeling → optimization	0,520	0,001	15,389
Multiobjective optimization → genetic algorithms, optimization	0,186	0,002	33,285
Multiobjective optimization → decision making, optimization	0,116	0,001	27,460
decision making, multiobjective optimization → optimization	0,500	0,001	14,797
Multiobjective optimization → energy efficiency, optimization	0,140	0,002	23,538
Energy efficiency, optimization → multiobjective optimization	0,257	0,002	23,538
Optimization → genetic algorithms, multiobjective optimization	0,060	0,002	15,783

Optimization → energy efficiency, genetic algorithms	0,048	0,002	16,538
Optimization → energy utilization, multiobjective optimization	0,043	0,001	14,374
Life cycle analysis → life cycle, optimization	0,070	0,001	31,833
Life cycle, optimization → life cycle analysis	0,577	0,001	31,833
Genetic algorithms, optimization → multiobjective optimization	0,364	0,002	33,285
Multiobjective optimization, optimization → genetic algorithms	0,453	0,002	29,060
Genetic algorithms, optimization → multiobjective optimization	0,364	0,002	33,285

Energy conservation

Energy savings → energy conversation	0,660	0,003	27,933
Energy conversation, energy utilization → energy consumption	0,179	0,002	27,100
Energy consumption → energy conversation, energy utilization	0,372	0,002	27,100
Energy utilization, information theory → building information modelling (BIM), energy conservation	0,395	0,001	86,450

Energy conservation, life cycle → energy utilization, life cycle assessment (LCA)	0,333	0,001	72,889
Building information modelling (BIM), energy utilization → energy conservation, information theory	0,137	0,001	64,754
Gas emissions → energy conservation, greenhouse gases	0,151	0,001	98,753
Greenhouse gases → energy conservation, gas emissions	0,078	0,001	61,569
Energy conservation, gas emissions → greenhouse gases	0,933	0,001	61,569
Energy savings, energy utilizion → energy conservation	0,884	0,002	36,551
Building energy consumption → energy conversation, energy utilization	0,392	0,002	1,623

Digital storage

Building information modelling (BIM), digital storage → information theory	0,483	0,002	5,541
Information theory → building information modelling (BIM), digital storage	0,028	0,002	5,541

Data handling			
Building information modelling (BIM), data handling → information theory	0,436	0,002	5,003
Data acquisition → data handling	0,122	0,001	11,250
Data handling → laser applications	0,102	0,001	11,104
Data mining			
Text mining → data mining	0,488	0,002	36,498
Data mining → text mining	0,133	0,002	36,498
Natural language processing systems → data mining	0,220	0,001	16,467
Data mining → big data	0,089	0,001	12,026
Forecasting			
Forecasting → predictive analytics	0.154	0.002	50.638
Predictive analytics → forecasting	0.519	0.002	50.638
Forecasting, neural networks → artificial neural networks	0,560	0,001	103,320

Artificial networks, forecasting → neural networks	0,778	0,001	78,496
Artificial neural networks → forecasting, neural networks	0,219	0,001	103,320
Forecasting → artificial neural networks, neural networks	0,089	0,001	23,779
Interoperability			
Data interoperability → interoperability	0,520	0,001	2,045
Building information modelling (BIM), interoperability → industry foundation classes (IFC), information theory	0.166	0.002	17,931
Industry foundation classes (IFC) → building information modelling (BIM), information theory, interoperability	0.075	0.002	13.710
Industry foundation classes (IFC), → building information modelling (BIM), interoperability	0.148	0.004	12,036
Building information modelling (BIM), industry foundation classes (IFC) → information theory, interoperability	0.086	0.002	14,937
Building information modelling (BIM), semantics → information theory, interoperability	0,096	0,001	16,637

Building information modelling (BIM), interoperability → information theory, semantics	0,110	0,001	15,698
Building information modelling (BIM), information theory, interoperability → industry foundation classes (IFC)	0.369	0.002	13.710
Information theory, interoperability → building information modelling (BIM), industry foundation classes (IFC)	0.353	0.002	14,937
Industry foundation classes (IFC)			
Industry foundation classes (IFC), information dissemination → information exchange	0,722	0,001	95,820
Information Exchange → Industry foundation classes (IFC), information dissemination	0,146	0,001	95,820
Industry foundation classes (IFC) → building information modelling (BIM), electronic data interchange	0.091	0.003	17.563
Building information modelling (BIM), industry foundation classes (IFC) → information theory, semantics	0.100	0.003	15.659
Building information modelling (BIM), interoperability → industry foundation classes (IFC), information theory	0.161	0.002	16.078

Building information modelling (BIM), semantics → industry foundation classes (IFC), information theory	0.186	0.003	20,109
Industry foundation classes (IFC), information theory → building information modelling (BIM), interoperability	0.213	0.002	16.078
Industry foundation classes (IFC), information theory → building information modelling (BIM), semantics	0.258	0.003	19.139
Industry foundation classes → building information modelling (BIM), information theory, semantics	0.087	0.003	14.608
Building information modelling (BIM), industry foundation classes (IFC) → information theory, interoperability	0.083	0.002	13.655
Industry foundation classes (IFC) → building information modelling (BIM), information theory, interoperability	0.072	0.002	12.300
Industry foundation classes (IFC) → building information modelling (BIM), semantics	0.182	0.005	13.465
Industry foundation classes (IFC) → Building information modelling (BIM), interoperability	0.155	0.005	16.078
Information theory, semantics → building information modelling (BIM), industry foundation classes (IFC)	0.404	0.003	15.659

Building information modelling (BIM), information theory, semantics → industry foundation classes (IFC)	0.434	0.003	14.608
Building information modelling (BIM), semantics → industry foundation classes (IFC)	0.400	0.005	13.465
Building information modelling (BIM), information dissemination → industry foundation classes (IFC)	00.389	0.002	13.091
Electronic data interchange → industry foundation classes (IFC)	0.368	0.003	12.402

Facility management

Building information modelling (BIM), office buildings → facility management (FM), information theory	0.172	0.003	59,878
Building information modelling (BIM), facility management (FM) → information theory, office buildings	0.309	0.003	41,500
Facility management (FM), information theory → building information modelling (BIM), office buildings	0.882	0.003	59,878
Facility management (FM), information theory → office buildings	0,971	0,003	29,462

Office buildings → building information modelling (BIM), facility management (FM), information theory	0.077	0.003	29,376
Building information modelling (BIM), facility management (FM), information theory → office buildings	0.968	0.003	29.376
Decision making, facility management (FM) → Office buildings	0,941	0,001	28,569
Facility management (FM), life cycle → building information modelling (BIM), Office buildings	0,583	0,001	39,86
Life cycle, office buildings → facility management (FM)	0.524	0.002	24.501
Building information modelling (BIM), Office buildings → facility management (FM), life cycle	0,080	0,001	39,586

"Optimization" keyword is one of the keywords with the highest lift value among the association rules of keywords determined as trends in the near future in node-based analysis. In cases where the multiobjective optimization keyword is present, the probability of encountering the genetic algorithms and optimization keywords was quite high. Also, when genetic algorithms and optimization keywords are seen together, the probability of seeing the multiobjective optimization keyword is very high. It was observed that multiobjective optimization, which is a multi-criteria decision-making mechanism with multiple objective functions, and genetic algorithm, which is an evolutionary optimization technique, are studied more than expected in optimization studies in the field of BIM. It can be easily said that multiobjective optimization studies are conducted together with genetic algorithm studies. In addition, studies, where optimization and genetic algorithm are studied together, focus on multiobjective optimization are trending in BIM data. It was observed that the energy conservation studies in the BIM field are studied more than expected, together with the areas of energy saving, energy consumption, and energy utilization. It was evidenced that energy saving studies are worked with energy conservation topics, energy conservation, energy consumption studies were combined with energy utilization topics, and energy consumption studies were combined with energy conservation and energy utilization topics more than expected. The probability of seeing the energy savings keyword and the energy conservation keyword together was quite high. Energy savings efforts are focusing on energy conservation issues. In cases where the energy conservation and energy consumption keywords are seen together, it is highly likely to see the energy usage keyword. Energy conservation and energy consumption studies focus on energy utilization are trending in BIM data. It was found more than expected in the research of digital storage studies in the field of BIM with the field of information theory, data mining and decision making were used together more than expected. The forecasting keyword, which was determined as a trend in the near future in node-based analysis, has the highest association rule lift value among other keywords determined as trends in the BIM data. Forecasting studies in the BIM field focus on predictive analytics are on the trending topics in BIM studies. It was observed that interoperability and information theory subjects in interoperability in the BIM field were more than expected in the BIM data. In cases where the interoperability and BIM keywords are seen together, it is highly likely that these two keywords can be seen together with the industry foundation classes (IFC) and information theory keywords. Trends in interoperability studies in the BIM field are focusing on industry foundation classes (IFC) and information theory. Also, in cases where the keywords building information modeling (BIM), information theory, and interoperability are seen together, the probability of seeing the industry foundation classes (IFC) keyword is quite high. For this reason, interoperability, and information theory studies in the field of BIM are focusing on industry foundation classes (IFC). In addition, it can be said that studies in the field of information exchange have started to focus on interoperability issues. In cases where the industry foundation classes (IFC) keyword is seen, the keywords BIM, semantics, information theory, interoperability, and electronic data exchange are highly likely to be seen. It has been observed that studies in the field of industry foundation classes (IFC) are more together than expected with BIM, semantics, information theory, interoperability, and electronic data interchange. Latest trends in BIM work with IFC are on information theory and semantics. In addition, it can easily be said that studies in the field of IFC focus on BIM and interoperability in the BIM data. In cases where the facility management (FM) keyword is seen, the life cycle and office buildings keywords are also likely to be seen. It is observed that the studies in the field of FM in BIM data are more than expected with the subjects of information theory and office buildings. The fields of information theory and office buildings in FM studies in BIM field are trending. In addition, life cycle topics are among the trends in FM studies in the BIM field.

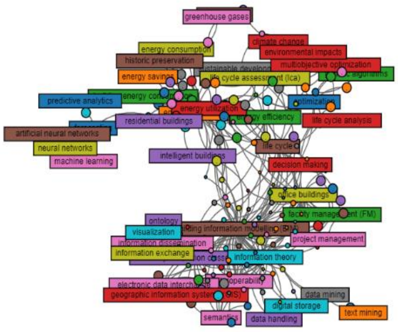


Figure 11. Visualization of Keywords Identified as Trends in the Node-based Analysis

The scatter plot helps us to more easily examine the association rules created by the desired keyword with other keywords. The scatter plot of the BIM keyword was shown in Figure 10. The confidence value gives the probability of the keywords appearing together in the data. The visualization of the rule graph created by choosing only the "Building Information Modelling" keyword is given in Figure 12.

The association rules with the highest confidence of the "building information modelling" keyword was given in Table 5. These rules, which have the highest confidence value and were created with building information modeling (BIM), were the subjects that are most likely to be seen together in studies in the research field of BIM. The *"Energy efficiency, energy utilization, information theory → building information modelling (BIM)"* association rule, which has the highest confidence value, was determined as the highest confidence value. This rule shows how important BIM adoption is for information management in the AECO-FM industry, such as energy efficiency and energy use. This rule was followed by the association rule "*Industry foundation classes (IFC), visualization → building information modelling (BIM)".* BIM plays a big role in the field of IFC and the visualization of structures. BIM, which provides great benefits in disseminating, processing, and analyzing information, is inevitable to be seen in IFC studies. BIM, which can keep up with technology rapidly, is also a catalyst in the adaptation of advanced technologies in the construction industry. Also, high confidence values of the rules show that the adoption of BIM in decision-making, and information theory studies play a significant role in the construction industry. The high confidence values obtained show that BIM plays a major role in different areas of the AECO-FM industry due to its features such as processing large data, providing control and supervision throughout the life cycle of the building, and hosting various data.

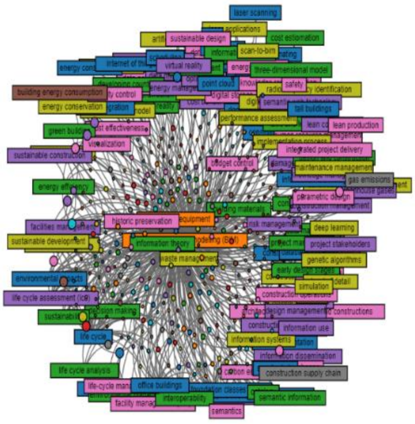


Figure 12. Rule Graph Created by Choosing Only the "Building information modelling" Keyword

Table 5. Association Rules with the Highest Confidence of the “Building Information Modelling” Keyword

Association rules	*Confidence*	*Support*	*Lift*
Information technology, information theory → building information modelling (BIM)	0,933	0,004	3,440
Decision making, information theory, life cycle, → building information modelling (BIM)	0,938	0,001	3,455
Construction management, information theory → building information modelling (BIM), project management	0,906	0,002	29,479

Information theory, semantics → building information modelling (BIM)	0,940	0,007	3,463
Information theory, office buildings → building information modelling (BIM)	0,932	0,007	3,434
Information theory, visualization → building information modelling(BIM)	0,918	0,004	3,385
Industry foundation classes (IFC), visualization → building information modelling (BIM)	0,947	0,002	3,491
Energy efficiency, energy utilization, information theory → building information modelling (BIM)	0,950	0,002	3,501
Geographic information systems (GIS), information systems, information theory → building information modelling (BIM)	0,944	0,001	3,481
Environmental impacts, information theory, life cycle assessment (LCA) → building information modelling (BIM), life cycle	0,941	0,001	39,691
Demolition, information theory → building information modelling (BIM)	0,933	0,001	3,440
Building information modelling (BIM),information systems, life cycle → information theory	0,929	0,001	10,645
Information theory, sustainability → building information modelling (BIM), sustainable development	0,926	0,002	68,763

Information theory, virtual reality→ building information modelling (BIM)	0,929	0,002	3,422
Industry foundation classes (IFC), office buildings → building information modelling (BIM)	0,931	0,002	3,431
Construction management, information theory → building information modelling (BIM)	0,935	0,002	3,448
Information theory, intelligent buildings → building information modelling (BIM)	0,932	0,003	3,434
Facility management (FM), information theory → building information modelling (BIM)	0,912	0,003	3,360
Facility management (FM), information theory → building information modelling (BIM), Office buildings	0,882	0,003	59,878
Facility management (FM), information theory, office buildings → building information modelling (BIM)	0,909	0,003	3,350
Information theory, life cycle, office buildings → building information modelling (BIM)	0,909	0,002	3,350
Augmented reality, information theory → building information modelling (BIM)	0,905	0,002	3,334
Information theory, scheduling → building information modelling (BIM)	0,900	0,002	3,317

Information theory, information use → building information modelling (BIM)	0,902	0,003	3,326
Information systems, information theory, information use → building information modelling (BIM)	0,875	0,001	3,225
Building information modelling (BIM), geographic information systems (GIS), information systems → information theory	0,895	0,001	10,257
Industry foundation classes (IFC), information dissemination → building information modelling (BIM)	0,889	0,001	3,276
Building information modelling (BIM), decision making, life cycle assessment (LCA) → life cycle	0,882	0,001	14,214
Information theory, risk assessment → building information theory (BIM)	0,880	0,002	3,243
Data integration, information theory → building information modelling (BIM)	0,870	0,002	3,205
Building information modelling (BIM), environmental impacts, life cycle assessment (LCA), sustainable development → life cycle	0,867	0,001	13,912
Building information modelling (BIM), environmental impacts, life cycle assessment (LCA) → life cycle	0,811	0,003	13,061

Information theory, laser applications → building information modelling (BIM)	0,857	0,002	3,159
Industry foundation classes (IFC), interoperability→ building information modelling (BIM)	0,855	0,004	3,149
Building information modelling (BIM), construction project management → project management	0,850	0,001	10,159
Historical preservation, information theory → building information modelling (BIM)	0,824	0,001	3,035
Information delivery → building information modelling (BIM)	0,818	0,002	3,015
Information systems, life cycle → building information modelling (BIM), information theory	0,812	0,001	9,850
Industry foundation classes (IFC) → building information modelling (BIM)	0,877	0,024	3,233

The support value indicates the frequency of an association rule in the data. The association rules with the highest support values of the "building information modelling" keyword are shown in Table 6. The rules of "*Building information modelling (BIM) → information theory*" (0.082*100 = 8.2%) and "*Information theory → building information modelling (BIM)*" (0.082*100 = 8.2%) with the highest support value was emerged as the most frequent association rules in BIM data. These values show that research field of BIM is mostly focused on providing information flow. These association rules were followed by "*Building information modelling (BIM) → project management*" and "*Project management*

→ *building information modelling (BIM)*". BIM provides a revolutionary platform for "design", "operations", "construction", "maintenance", and management of constructions or built environment. BIM plays a major role in project management studies as it improves communication and coordination in the AECO-FM industry. BIM, which plays a major role in energy management in the AECO-FM industry, is frequently involved in sustainable development issues. In addition to these, life cycle, decision-making, and visualization issues are also seen quite frequently in the BIM field. BIM, which is involved in the entire life cycle of structures with its various tools, takes a great place in the decision-making stages. In addition, being compatible with the IFC format, is very convenient for visualization and monitoring. Moreover, it is observed that there are many studies in the field of BIM and semantics, which also includes non-geometric dimensions of the BIM model.

Table 6. Association Rules with the Highest Support Values of the "Building Information Modelling" Keyword

Association rules	*Confidence*	*Support*	*Lift*
Building information modelling (BIM) → information theory	0,304	0,082	3,485
Information theory → building information modelling (BIM)	0,946	0,082	3,485
Industry foundation classes (IFC) → building information modelling (BIM)	0,877	0,024	3,233
Project management → building information modelling (BIM)	0,367	0,031	1,354

Building information modelling (BIM) → project management	0,113	0,031	1,354
Building information modelling (BIM) → life cycle	0,087	0,024	1,408
Building information modelling (BIM) → decision making	0,084	0,023	1,103
Building information modelling (BIM) → information theory, Project management	0,050	0,014	3,618
Information theory, Project management → building information modelling (BIM)	0,982	0,014	3,618
Information theory → building information modelling (BIM), decision making	0,116	0,010	5,071
Building information modelling (BIM) → semantics	0,052	0,014	1,966
Semantics → building information modelling (BIM)	0,534	0,014	1,966
Sustainable development → building information modelling (BIM)	0,222	0,013	0,817
Building information modelling (BIM), project management → information theory	0,441	0,014	5,053
Decision making, information theory → building information modelling (BIM)	0,937	0,010	3,453

Visualization → building information modelling (BIM)	0,463	0,011	1,707
Building information modelling (BIM), life cycle → information theory	0,479	0,011	5,486
Building information modelling (BIM) → information theory , life cycle	0,042	0,011	3,527
Life cycle → building information modelling (BIM), information theory	0,183	0,011	2,216

The association rules with higher lift values than the others, represented by dark dots in the BIM scatterplot, are listed in Table 7. *"Information theory, life cycle assessment (LCA) → building information modelling (BIM), environmental impact, life cycle"* and *"building information modelling (BIM), environmental impact, life cycle → information theory, life cycle assessment (LCA)"* are determined as association rules with the highest lift values. Information theory and life cycle assessment (LCA) studies are focusing on BIM and environmental impact. It is obvious that BIM, which stores, manages, and processes data throughout the life cycle of structures, are effective in environmental impacts estimation. After these association rules, "*Building information modelling (BIM), cost estimating → cost estimation"* were determined as association rules with the highest lift values. Studies in the field of BIM are concentrating on the field of construction estimation. These association rules are followed by the rules created by the building information modeling (BIM) keyword with the lean production and lean construction keywords. In cases where the word lean construction is seen, the probability of encountering the building information modeling (BIM) and lean production keywords is much higher than expected. In addition, in cases where the lean production keyword is seen, the probability of encountering building information modeling (BIM) and lean construction keywords is much higher than expected. In cases where the building information modeling (BIM) and lean construction keywords are together, the probability of seeing the lean production keyword is higher than expected. Lean construction research in the BIM field will be one of the trending topics in the BIM data. In addition, it can easily be said that lean construction studies in the BIM field are the focus of lean production in BIM data. The lift value of the association rules between the keywords building information modeling (BIM), intelligent buildings, sustainable development, and information theory are also quite high. In cases where the words information theory and sustainable development are seen together, the probability of encountering building information modeling (BIM) and intelligent keywords is much higher than expected. In studies in the field of BIM, sustainability development and information theory are included in the trends in the BIM studies. BIM, which provides great benefits to stakeholders in data flow and information management, also play a significant role in developments in the field of sustainability. The fact that BIM contains various data and provides great benefits to users in the processing of this data has made BIM a center in the field of sustainability. It also plays a catalyst role on the Internet of Things (IoT) studies in the construction industry, as it adapts very easily to new technologies. In addition, it can be said that intelligent buildings are the focal point on sustainable development studies lately. In recent years, issues such as environmental pollution, greenhouse gases, consumption of energy resources have begun to take place on the world agenda. It is obvious that topics such as energy use and energy efficiency studies in the AECO-FM industry are among the trends with BIM and information theory. In addition to these association rules, the association rules created by the building information modeling (BIM) keyword with the keywords facility management (FM), information theory, and office buildings have very high lift values. In cases where the words facility management (FM) and information theory are seen together, the probability of encountering the keywords building information modeling (BIM) and office buildings is higher than expected. BIM, which also enables cost, energy, and risk analysis, provides interoperability and collaboration between stakeholders as it can be applied to the entire life cycle. It can be said that research in which the subjects of facility management (FM) and information theory are studied together are among the trends in the BIM studies. Geographic information Systems (GIS) enables the collection, storage, processing, management, spatial analysis, querying, and presentation of large-scale geographic data. Adaptation of GIS to the AECO-FM industry is of great benefit in studies on large areas. BIM, which specializes in processing data in different formats, is accelerating this adaptation. For this reason, BIM and GIS studies are included in the trends in the BIM studies. BIM allows the use of geometric and non-geometric data in different formats. In addition, its ability to work in harmony with IFC makes it much easier for stakeholders to share information. It is obtained from these results that information exchange, information dissemination, and building information modeling (BIM) keywords are seen together more than expected in studies in the field of BIM. It can be easily said that information exchange and information dissemination are among the trends in the field of BIM. Due to the increasing large structures and the difficulty of processing data, studies in the field of data visualization have been increasing in recent years (Cheng and Teizer 2013; Guo, He et al., 2017). Adaptations of developments in data visualization to BIM are among the trend topics in the BIM field. The rules created by the laser scanning keyword with BIM have high lift values. Providing high accuracy, laser scanning applications provide 3D precise shape and size mapping of objects by using point cloud system in built environments and construction sites. In addition to these, semantics, accident prevention, demolition, safety, supply chain, gas emissions, and greenhouse gases keywords also have high lift values in the rules they created with BIM. BIM plays a huge role in the AECO-FM industry due to its compatibility with technology, accompanying stakeholders throughout the lifecycles of structures, with its ability to store, share and process data, and many other benefits. For this reason, in the face of the challenges and challenges faced in the AECO-FM industry, the integration of BIM into emerging technologies and the adaptation of BIM not only to design and construction but also to further phases of the life cycle are inevitable.

Most of the lift values of association rules of recent trends detected in the node-based scientometric analysis could not reach the highest values among association rules determined in rule-based analysis. Among the highest lift values of the two analyses, the lowest lift value is 2.359. For this reason, the lift value limit in the table is increased from 0.007 to 2.000, the code was run again, and the knowledge graph is visualized again in Figure 13. With the visualized graphics, keywords and association rules that have been studied more than expected in the BIM field were made easy to examine. These association rules, which are seen together more than expected, actually represent trends in the BIM studies. The trends in the research field of BIM are discussed in depth in the following section.

Table 7. List of Association Rules with the Highest Lift Value of the "Building Information Modelling" Keyword

Association rules	*Confidence*	*Support*	*Lift*
Building information modelling (BIM), lean production → lean construction	0,750	0,002	67,603
Lean construction →building information modelling (BIM), lean production	0,160	0,002	67,603
Building information modelling (BIM), lean construction → lean production	0,467	0,002	65,600
Lean production → building information modelling (BIM), lean construction	0,250	0,002	65,600
Facility management (FM), information theory → building	0.882	0.003	59,878

information modelling (BIM), office buildings			
Building information modelling (BIM), office buildings → facility management (FM), information theory	0,172	0,003	59,878
Facility management (FM) → building information modelling (BIM), office buildings	0,464	0,006	31,492
Facility management (FM) → building information modelling (BIM), life cycle, office buildings	0,092	0,001	32,742
Building information modelling (BIM), office buildings → Facility management (FM)	0,408	0,006	31,492
Building information modelling (BIM), life cycle office buildings → Facility management (FM)	0,424	0,001	32,742
Information theory, life cycle assessment (LCA) → building information modelling (BIM), life cycle	0.963	0.002	40.610
Environmental impacts, information theory, life cycle → building information modelling (BIM), life cycle assessment (LCA)	0,762	0,001	115,341

Building information modelling (BIM), life cycle assessment (LCA) → Environmental impacts, information theory, life cycle	0,205	0,001	115,341
Building information modelling (BIM), life cycle → information theory, life cycle assessment (LCA)	0,093	0,002	40,610
Environmental impacts, information theory, life cycle assessment (LCA) → building information modelling (BIM), life cycle	0,941	0,001	39,691
Information theory, sustainability → building information modelling (BIM), sustainable development	0,926	0,002	68,763
Building information modelling (BIM), sustainable development → information theory, sustainability	0,157	0,002	68,763
Energy conservation, information theory → building information modelling (BIM), energy utilization	0,680	0,001	64,754
Building information modelling (BIM), geographic information systems (GIS), information theory → information systems	0,654	0,001	41,960
Geographic information systems (GIS), information theory → building	0,630	0,001	81,700

information modelling (BIM), information systems			
Building information modelling (BIM), information systems → geographic information systems (GIS), information theory	0,187	0,001	81,700
Information theory, life cycle, life cycle assessment (LCA) → building information modelling (BIM), environmental impacts	0,615	0,001	90,831
Building information modelling (BIM), environmental impacts → information theory, life cycle, life cycle assessment (LCA)	0,200	0,001	90,831
Building information modelling (BIM), information dissemination → information exchange	0,600	0,002	79,604
Information exchange → building information modelling (BIM), information dissemination	0,303	0,002	79,604
Energy conservation, information theory → building information modelling (BIM), energy utilization	0,680	0,001	64,754
Building information modelling (BIM), energy utilization → energy conservation, information theory	0,137	0,001	64,754

Energy utilization, information theory → building information modelling (BIM), energy efficiency	0,442	0,002	32,407
Building information modelling (BIM), energy efficiency → energy utilization, information theory	0,118	0,002	32,407
Building information modelling (BIM), geographic information systems (GIS), information theory → information systems	0,654	0,001	41,960
Information systems → building information modelling (BIM), geographic information systems (GIS), information theory	0,092	0,001	41,960
Information theory, life cycle assessment (LCA) → building information modelling (BIM), environmental impact, life cycle	0,593	0,001	166,603
Building information modelling (BIM), environmental impact, life cycle → information theory, life cycle assessment (LCA)	0,381	0,001	166,603
Building information modelling (BIM), life cycle assessment (LCA), sustainable development → environmental impacts, life cycle	0,591	0,001	43,609

Building information modelling (BIM), life cycle → environmental impacts, information theory, life cycle assessment (LCA),	0,057	0,001	39,691
Environmental impacts, life cycle → building information modelling (BIM), life cycle assessment (LCA), sustainable development	0,081	0,001	43,609
Environmental impacts, life cycle → Building information modelling (BIM), information theory, life cycle assessment (LCA)	0,100	0,001	43,733
Facility management (FM), life cycle → building information modelling (BIM), office buildings	0,583	0,001	39,586
Building information modelling (BIM), office buildings → facility management (FM), life cycle	0,080	0,001	39,586
Information theory, intelligent buildings → building information modelling (BIM), sustainability development	0,523	0,002	38,820
Building information modelling (BIM), sustainability development → information theory, intelligent buildings	0,145	0,002	38,820

Environmental impacts, information theory → building information modelling (BIM), life cycle assessment (LCA)	0,531	0,001	80,423
Building information modelling (BIM), life cycle assessment (LCA) → environmental impacts, information theory	0,218	0,001	80,423
Environmental impacts, information theory → building information modelling (BIM), life cycle, life cycle assessment (LCA)	0,500	0,001	123,000
Environmental impacts, information theory → building information modelling (BIM), sustainable development	0,438	0,001	32,491
Building information modelling (BIM), sustainable development → environmental impacts, information theory	0,088	0,001	32,491
Building information modelling (BIM), environmental impacts life cycle assessment (LCA) → information theory, life cycle	0,432	0,001	36,473
Information theory, life cycle → building information modelling	0,114	0,001	36,473

(BIM), environmental impacts life cycle assessment (LCA)			
Data visualization, information theory → building information modelling (BIM), visualization	0,483	0,001	45,241
Building information modelling (BIM), visualization → data visualization, information theory	0,111	0,001	45,241
Information theory, visualization → → building information modelling (BIM), data visualization	0,286	0,001	63,655
Construction and demolition waste → building information modelling (BIM), demolition	0,138	0,001	58,322
Building information modelling (BIM), demolition → construction and demolition waste	0,464	0,001	58,322
Building information modelling (BIM), laser scanning → laser applications	0,447	0,001	48,912
Laser applications → building information modelling (BIM), laser scanning	0,157	0,001	48,912

Accident prevention, building information modelling (BIM) → safety	0,452	0,001	42,323
Accident prevention → building information modelling (BIM), safety	0,095	0,001	37,486
Building information modelling (BIM), information exchange → information dissemination	0,422	0,002	68,240
Information dissemination → building information modelling (BIM), information exchange	0,370	0,002	68,240
Energy utilization, information theory → building information modelling (BIM), energy conservation	0,395	0,001	86,450
Information theory, sustainable development → building information modelling (BIM), sustainability	0,357	0,002	34,286
Building information modelling (BIM), sustainability → information theory, sustainable development	0,203	0,002	34,286
Information theory, office buildings → building information modelling (BIM),facility management (FM)	0,341	0,003	41,500

Building information modelling (BIM),facility management (FM) → information theory, office	0,309	0,003	41,500
Energy efficiency, information theory → building information modelling (BIM),energy utilization	0,339	0,002	32,309
Building information modelling (BIM),energy utilization → energy efficiency, information theory	0,153	0,002	32,309
Information theory, sustainable development → building information modelling (BIM), intelligent buildings	0,329	0,002	48,497
Building information modelling (BIM), environmental impacts, life cycle → life cycle assessment (LCA), sustainable development	0,310	0,001	45,686
Building information modelling (BIM), cost estimating → cost estimation	0,292	0,001	164,000
Cost estimating →building information modelling (BIM), cost estimation	0,171	0,001	134,400
Construction supply chain → building information modelling (BIM), supply chain	0,275	0,001	66,151

Supply chain → → building information modelling (BIM), construction supply chain	0,074	0,001	54,667
Building information modelling (BIM), information use → information systems, information theory	0,250	0,001	52,714
Laser scanning → → building information modelling (BIM), laser applications	0,207	0,001	58,286
Gas emissions → building information modelling (BIM), greenhouse gases	0,183	0,001	67,452
Greenhouse gases → building information modelling (BIM), gas emissions	0,095	0,001	65,966
Building information modelling (BIM), geographic information systems (GIS) → information systems, information theory	0,179	0,001	37,732
Building information modelling (BIM), information systems → information theory, information use	0,154	0,001	44,308
Demolition → Building information modelling (BIM), waste management	0,116	0,001	42,694

Building information modelling (BIM), semantics → semantic information	0,090	0,001	39,281
Semantics→ building information modelling (BIM), semantic information	0,048	0,001	35,367

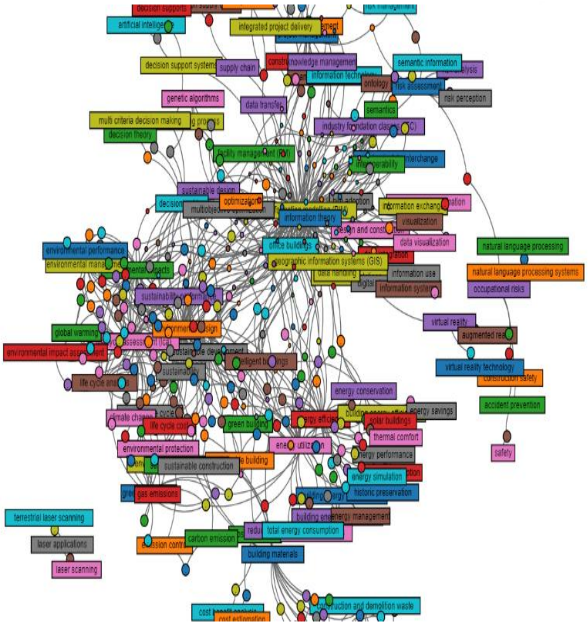


Figure 13. Knowledge Graph Visualization of BIM Data with Lift Limitation

3.6.Trend prediction by Utilizing Machine Learning Method

The lift values of some of the rules of the trend topics determined in the BIM data in 2013-2021 were exported with the Apriori algorithm. The regression values of these data were taken, and it was determined whether they would be among the trends in the near future. Random forest algorithm was chosen for regression. This algorithm uses ensemble learning method for regression. Thus, the algorithm combines multiple machine learning algorithms to make a more accurate prediction than a single model. The lift values of the rules by years, and regression trend prediction values of 2021 and 2023 are given in the Table 8 and Table 9.

Table 8. Lift Values of the Trend Topics Determined in the BIM Data by Years

Association Rules	Lift Values by Years							
	2013	2014	2015	2016	2017	2018	2019	2020
Building information modelling → decision making	0,844	1,474	1,310	1,148	1,132	1,296	1,658	1,191
Building information modelling → industry foundation classes	2,302	1,778	1,468	1,635	2,648	2,374		
Building information modelling → life cycle	2,110	1,500	0,881	1,417	2,257	1,553	1,606	1,740
Building information modelling→ facility management	1,876	1,818		2,559		2,481	3,503	
Environmental impacts → life cycle					11,725	8,260	9,064	7,159

As an example, the lift values of the "*Building information modeling → decision making*" rule which passed the threshold set in the data of all years and usually had lift values above 1 between 2013-2020 were determined and shown on Figure 14. As a result of the regression of these values, the expected value for 2021 was determined as 1,331 and the expected value for 2023 was determined as 1,364. The lift value for 2021 was determined as 1,435. These rules, which are determined as a trend in BIM data, will be among the trends in the near future by having lift values above 1.

Table 9. Regression Trend Prediction Values and Determined Lift Value of 2021 of The Trend Topics Determined in The BIM Data by Years

Association Rules	Regression Trend Prediction Value for 2021	Lift Value for 2021	Regression Trend Prediction Value for 2023
Building information modelling → decision making	1,331	1,435	1,364
Building information modelling → industry foundation classes	2,255		2,338
Building information modelling → life cycle	1,716	2,046	1,705
Building information modelling→ facility management	2,459		2,542
Environmental impacts → life cycle	7,764	7,336	7,758

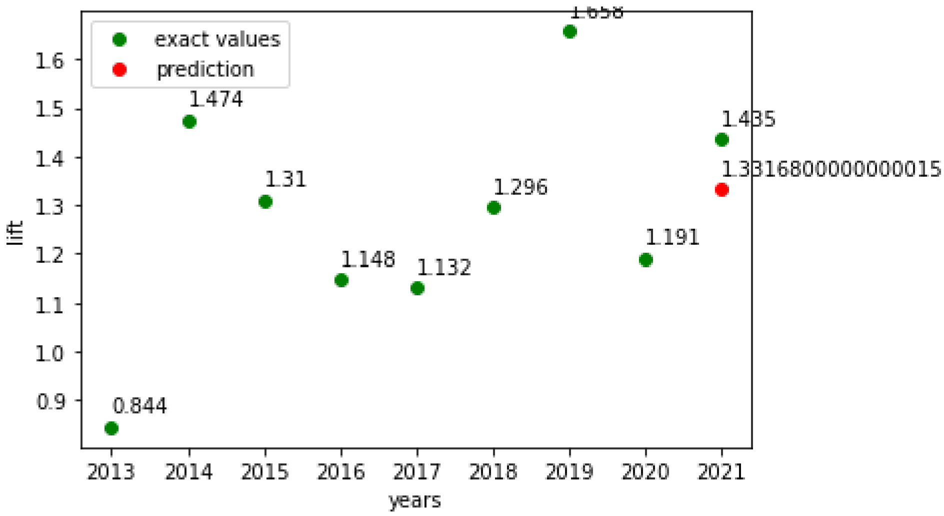


Figure 14. Lift values and regression of the "building information modelling → decision making" rule by years

"*Building information modeling → life cycle*" association rule is another rule that could exceed the threshold and have lift values above 1 in all years. As a result of the regression of these values, the predicted value for 2021 was determined as 1.716 and the predicted value for 2023 was determined as 1.705. The lift value for 2021 was determined as 2,046.

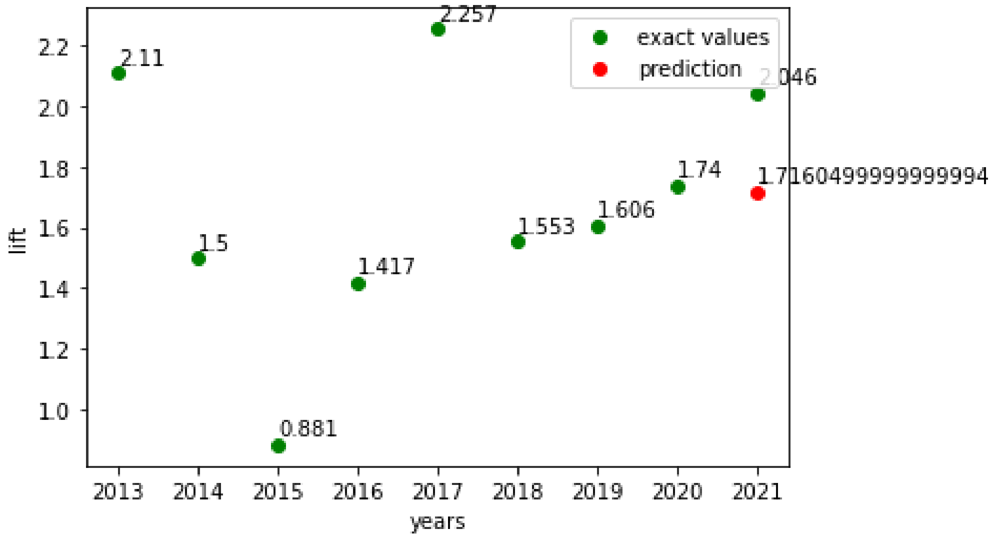


Figure 15. Lift values and regression of the “building information modelling → life cycle” rule by years

The estimated lift values of the association rules in Table 9 for 2023 are all above 1. For this reason, it can easily be said that these rules will be encountered in studies in the field of BIM more than expected in 2023. In addition, considering the rules of association, it can be easily said that studies in the field of BIM will focus on "decision making", "life cycle", "facility management", "industry foundation classes", and "environmental impacts" and these topics will be among the trends in 2023.

4.DISCUSSION

4.1.The Future Trends in BIM Research

Although the BIM literature initially focused on nD designs, as technology developed, and systems, have been rapidly adopted due to the ease of information management. However, the increase in different methods and tools thanks to the rapidly developing technology has increased the need for studies in the fields of information dissemination and information developed, BIM tools were integrated to the platform. Thus, sustainability studies in the literature have begun to focus on the more effective use of information accumulated in Building Information Models. Using the data BIM collects throughout the life cycle of buildings enables more sustainable, economical, and safer design, operation, and maintenance. Therefore, studies in the near future tend to be related to the fields of optimization, life cycle assessment (LCA), intelligent buildings, energy utilization, and energy conservation.

As it is known, the adoption studies in the field of BIM are mostly focused on architectural design, building information model, information management, and information theory. Building information models, which offer multidimensional solutions in architectural design and information management exchange. For this reason, studies in the field of interoperability and IFC have gained momentum recently. In addition to the recent increase in studies on these topics, adoption-based BIM studies will also show a trend in interoperability and IFC in the near future.

BIM offering the nD solutions has undoubtedly become an indispensable approach to project management in the AECO-FM industry. BIM, which offers great benefits to stakeholders in the decision-making phases by hosting, processing, and making data meaningful, offers great benefits not only in the design phase but also in the operations, facilities management, and demolition phases. BIM, which is predictive, also plays an important role in the risk assessment of construction projects. Management studies in the field of BIM particularly focus on construction management, project management, and risk assessment. Due to the rapid adaptation of BIM to evolving technology and the widespread use of BIM in the management of life cycles of buildings, the field of Facility Management studies have gained momentum recently.

The increase in complex structures and the difficulty of collecting information in various areas such as energy efficiency, cost control, and safety caused for need in various optimization and decision-making mechanisms for BIM tools. The processing of information collected in building information models by decision support system tools has been of great help in forecasting in building life cycle. By accelerating the integration of multi-purpose optimization tools in BIM, stakeholders can be presented with different alternatives by using the information collected in the building information models. In particular, studies in the field of multiobjective optimization which contains multiple functions and serves multiple purposes, and the genetic algorithms based on the natural selection process that mimics biological evolution have increased rapidly in recent studies. However, more detailed research is needed to improve the performance of capturing, understanding, processing, and decision-making in a single environment for all the information collected in different building information models during the life of the building. Therefore, there is a need for advanced integration of BIM and data mining and deep learning techniques. For this reason, studies in the fields of data handling, data mining, deep learning, digital storage, and forecasting are increasing. However, there is a need for studies on adapting systems such as the Internet of Things (IoT), laser scanning, GIS, and point cloud to meet data needs through the regular data flow.

4.2.A Novel Trend Prediction System Design Via Association Rule Mining Represented in Knowledge Graphs: BIM Research Field Application

In the first phase of the study, the data containing 12185 articles in Building Information Modeling (BIM) were analyzed using the traditional scientometric analysis tool and the analysis results were discussed. Since this study was conducted using the node-based analysis tool, only the links between two keywords and the keywords divided into groups by the classification method could be compared with the keywords of the group they formed. Moreover, since keywords with the same meaning could not be combined, the keywords with the same meaning that were important for the data had to be selected with different spellings for the analysis. This thesis presents some proposed solutions to overcome the limitations of the traditional scientometric analysis tools. First of all, the rule-based analysis approach was preferred as the analysis method instead of the node-based analysis approach. In this study, the Apriori algorithm, which determines the relationships between items and groups of items, was used as the association rule mining algorithm. The Apriori algorithm identifies items, recursively selects frequently recurring item groups that are above the minimum threshold set by the user and expands these item groups by performing as many repetitions as possible within the set threshold. In this way, the Apriori algorithm allows the user to determine the desired threshold and explore all relationships between item groups. The Apriori algorithm does not stop the analysis at the first iteration. After finding the binary relations, it finds the triple and multiple relations by iterating through the phases of merging and pruning. This recursive mining approach enables the exploration of deeper and hidden relationships. In addition, the pruning phase ensures that the relationships that do not reach the threshold are pruned in each iteration, allowing the algorithm to work more effectively. The user is presented with a visualized knowledge graph that provides the ability to define all items up to the specified threshold and examine frequently recurring item groups. Moreover, unlike node-based analysis, it is possible to observe not only the relationship between two items but also the relationship between multiple items. The threshold value for multiple variables can be defined by the user. The user can make changes to the visualization by selecting or excluding preferred keywords. For this purpose, users do not have to start the analysis from the beginning again. In addition, synonymous keywords that are important to the data found in the previous study, equalized to each other in the data preprocessing section, and their support, confidence, and lift values were maintained. Thus, the analysis results provided more accurate and precise results compared to the traditional scientometric analysis results. In addition, it was made easier for the researcher to interpret the results of the analysis and draw conclusions by presenting clear and quantitative values for a robust assessment. Besides, this study presented a dynamic estimation approach apart from the static results of other traditional scientometric analysis approaches. made a trend prediction for the near future studies by monitoring the changes in the data with the machine learning approach and using multiple machine learning techniques.

4.3. Validation

The lift value indicates how much more or less an association rule is in the data than expected. The lift value of the "Building information modeling → decision making" rule discussed in this study has generally remained above 1 over the years. The results of the estimation system presented in this study also predicted the lift value above 1 during the year interval determined for this association rule. The frequency of occurrence of keywords that make up association rules in the data and the lift value, which depends on the frequency of occurrence together in the data, is directly proportional to the probability of finding the rule they represent in the data. For this reason, the frequent occurrence of the keywords that make up the association rule examined in the data will directly affect the lift value. For this reason, the number of articles containing the keywords "building information modeling" and "decision making" in the BIM data and the change in the lift values presented in this study over the years are shown in Figure 16. As seen in Figure 16, the estimated lift values of the association rule and the number of articles in which the keywords represented by the association rule increase over the years are increasing in the same proportion. The increase in the lift value depends on the increase in the confidence value in the association rule it represents. In this case, it is concluded that the increase in the probability of both keywords being together in the data increases the lift value. The increase in the number of articles with both keywords and the lift value in the same proportion proves the accuracy of the method adopted by this study.

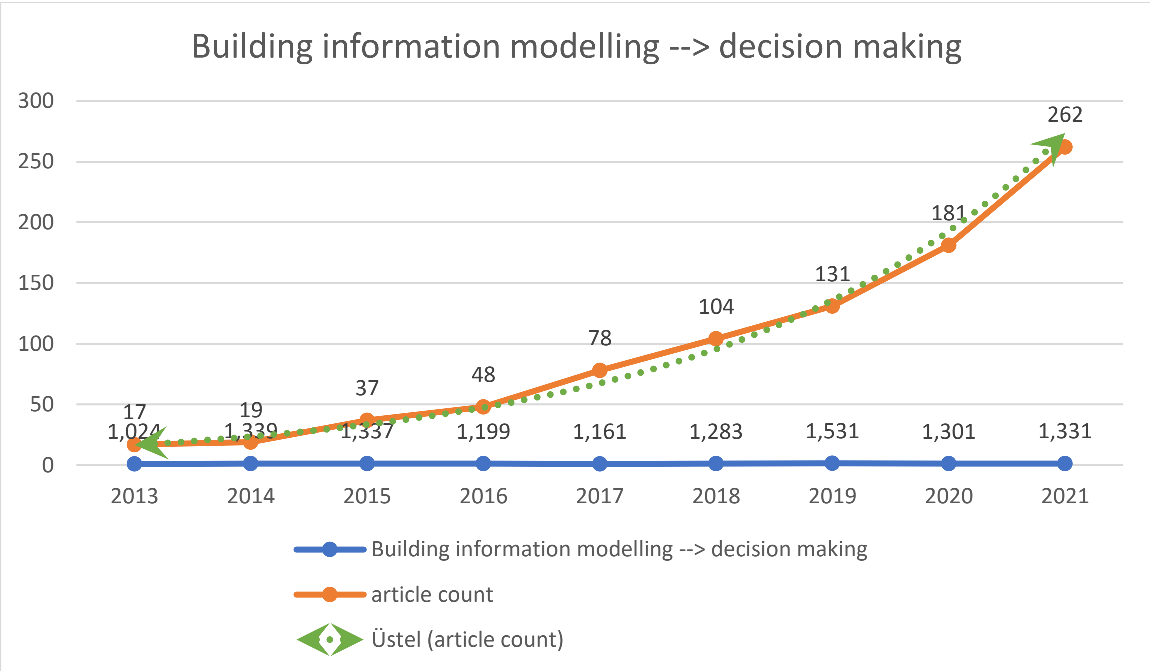


Figure 16. Building information modelling → decision making association rule and article counts by years

4.4.Limitations of the Thesis

The Apriori algorithm incrementally expands the item groups by analyzing all item groups that meet or exceed the threshold set by the researcher. However, this iterative process can take a long time to obtain analysis results when the threshold is very low, and the amount of data is very large. In this study, it was not necessary to set the threshold too low. But in future studies, if you work with big data or set the threshold too low, it will take longer to get the analysis results. Synonymous words that are important to the BIM data analyzed in this study are synchronized with each other during the preprocessing of the data, so data loss is avoided as much as possible. However, it is exceedingly difficult for the researcher to manually equate all synonymous words in the data. In addition, the Apriori algorithm provides the researcher with values for Support, Lift, and Confidence. By examining these values, the dominance of items and item groups in the data, the strength of relationships between items, and the frequency with which item groups occur more frequently than expected can be examined. However, there is no information on citations and publication years for the items examined in this study. In addition, only keyword analysis can be conducted in this study. This study considered rules with confidence values of 0.95 and above as parasites and these rules are not included in the analysis. The reason for this is that the keywords that create these rules are so close to each other that the library counts their confidence value as 1. In this developed system, confidence, support, and lift values are drawn manually. But it can be very difficult to manually export analysis results containing thousands of rules. Besides, the knowledge graph developed in this study is static. It does not give researchers dynamic results.

4.5.Suggestions for Future Studies

In future studies, various combinations of hybrid algorithms can be used to work with larger data more easily and quickly. Also in future studies, research can be conducted so that synonymous words can be easily combined by the researcher. In this study, a word analysis was carried out by working only on the values of support, confidence, and lift. In future studies, the knowledge graph can be enriched by adding citation and average publication year values to the analysis. In addition to these, various analysis types such as author and document citation analyzes can be developed. To eliminate confidence values above 0.95, which are considered parasites, from the analysis, various solutions can be proposed for this problem in future studies. In addition, time values can be added to the analysis and the knowledge graph can be made dynamic. With the time dimension, the dynamism of the knowledge graph in the process can be examined. In this study, confidence, support, and lift values were drawn manually. In future studies, this problem can be solved by developing an interface for users to export values automatically. In future studies, the source code can be developed by taking these limitations into account.

5.CONCLUSION

Recently, various analysis methods have been used to examine the rapidly increasing data with the development of technology. In addition, analysis results can be visualized with various tools, allowing the user to make comments more easily. Solutions were provided for the traditional scientometric analysis tools' limitations encountered in the BIM research as a case study and were discussed in this thesis. A knowledge graph was created from the data of the BIM research field using the Apriori Algorithm. A scientometric analysis was carried out by performing a rule-based analysis. Thanks to the capabilities of the newly designed system for scientometric analysis, the relevance and support of the items and groups of items for the research field could be investigated. Moreover, the relationships between multiple items were made observable, and the dominance of these item groups in the data can also be observed. The effectiveness of the important synonymous keywords equated to each other in the data was more accurately determined in the new analyzing tool. In future studies, the knowledge graph can be developed to be dynamic by adding a time dimension to the analysis. In addition, the analysis results can be enriched by adding various variables and features to the analysis. In addition, an interface can be created for researchers to combine their preferred synonyms more easily.

6.REFERENCES

1.Volk, Rebekka, Julian Stengel, and Frank Schultmann. "Building Information Modeling (BIM) for existing buildings—Literature review and future needs." *Automation in Construction,* 38 (2014): 109-127.

2.Bryde, David, Martí Broquetas, and Jürgen Marc Volm. "The project benefits of building information modelling (BIM)." *International Journal of Project Management*, 31.7 (2013): 971-980.

3.Succar, Bilal. "Building information modelling framework: A research and delivery foundation for industry stakeholders." *Automation in Construction*, 18.3 (2009): 357-375.

4.Azhar, Salman. "Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry." *Leadership And Management in Engineering*,11.3 (2011): 241-252.

5.Van Eck, Nees, and Ludo Waltman. "Software survey: VOSviewer, a computer program for bibliometric mapping." *Scientometrics*, 84.2 (2010): 523-538.

6.Chen, Chaomei. "The citespace manual." *College of Computing and Informatics,* 1 (2014): 1-84.

7.Batagelj, Vladimir, and Andrej Mrvar. "Pajek—analysis and visualization of large networks." Graph drawing software. *Springer*, Berlin, Heidelberg, 2004. 77-103.

8.Siebert, Sebastian, and Jochen Teizer. "Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system." *Automation in Construction*, 41 (2014): 1-14.

9.Gheisari, Masoud, and Behzad Esmaeili. "Applications and requirements of unmanned aerial systems (UASs) for construction safety." *Safety Science,* 118 (2019): 230-240.

10.De Melo, R. R. S., Costa, D. B., Álvares, J. S., & Irizarry, J. (2017). Applicability of unmanned aerial system (UAS) for safety inspection on construction sites. *Safety Science*, 98, 174-185.

11.Darko, A., Chan, A. P., Adabre, M. A., Edwards, D. J., Hosseini, M. R., & Ameyaw, E. E. (2020). Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation in Construction*, 112, 103081.

12.He, Q., Wang, G., Luo, L., Shi, Q., Xie, J., & Meng, X. (2017). Mapping the managerial areas of Building Information Modeling (BIM) using scientometric analysis. *International Journal of Project Management*, 35(4), 670-685.

13.Kalinichuk, S. (2015). Building information modeling comprehensive overview. *Journal of Systems Integration*, 6(3), 25.

14.Chatzimichailidou, M., & Ma, Y. (2022). Using BIM in the safety risk management of modular construction. *Safety Science*, 154, 105852.

15.Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., ... & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141, 104440.

16.Pu, H., Fan, X., Schonfeld, P., Li, W., Zhang, W., Wei, F., ... & Li, C. (2022). Extending IFC for multi-component subgrade modeling in a railway station. *Automation in Construction*, 141, 104433.

17.Zhai, R., Zou, J., He, Y., & Meng, L. (2022). BIM-driven data augmentation method for semantic segmentation in superpoint-based deep learning network. *Automation in Construction*, 140, 104373.

18.John Samuel, I., Salem, O., & He, S. (2022). Defect-oriented supportive bridge inspection system featuring building information modeling and augmented reality. *Innovative Infrastructure Solutions*, 7(4), 1-17.

19.Han, B., & Leite, F. (2022). Generic extended reality and integrated development for visualization applications in architecture, engineering, and construction. *Automation in Construction*, 140, 104329.

20.Bellagarda, A., Cesari, S., Aliberti, A., Ugliotti, F., Bottaccioli, L., Macii, E., & Patti, E. (2022). Effectiveness of neural networks and transfer learning for indoor air-temperature forecasting. *Automation in Construction*, 140, 104314.

21.Zavari, M., Shahhosseini, V., Ardeshir, A., & Sebt, M. H. (2022). Multi-objective optimization of dynamic construction site layout using BIM and GIS. *Journal of Building Engineering*, 52, 104518.

22.Wang, W., Guo, H., Li, X., Tang, S., Li, Y., Xie, L., & Lv, Z. (2022). BIM Information Integration Based VR Modeling in Digital Twins in Industry 5.0. *Journal of Industrial Information Integration*, 28, 100351.

23.Zhang, F., Chan, A. P., Darko, A., Chen, Z., & Li, D. (2022). Integrated applications of building information modeling and artificial intelligence techniques in the AEC/FM industry. *Automation in Construction*, 139, 104289.

24.Nassereddine, H., Hanna, A. S., Veeramani, D., & Lotfallah, W. (2022). Augmented Reality in the Construction Industry: Use-Cases, Benefits, Obstacles, and Future Trends. *Frontiers in Built Environment*, 8, 730094.

25.Shen, W., Hao, Q., Mak, H., Neelamkavil, J., Xie, H., Dickinson, J., ... & Xue, H. (2010). Systems integration and collaboration in architecture, engineering, construction, and facilities management: A review. *Advanced Engineering Informatics*, 24(2), 196-207.

26.Singh, Vishal, Ning Gu, and Xiangyu Wang. "A theoretical framework of a BIM-based multi-disciplinary collaboration platform." *Automation in Construction*, 20.2 (2011): 134-144.

27.Ghaffarianhoseini, A., Tookey, J., Ghaffarianhoseini, A., Naismith, N., Azhar, S., Efimova, O., & Raahemifar, K. (2017). Building Information Modelling (BIM) uptake: Clear benefits, understanding its implementation, risks and challenges. *Renewable and Sustainable Energy Reviews*, 75, 1046-1053.

28.Ding, L. Y., Zhong, B. T., Wu, S., & Luo, H. B. (2016). Construction risk knowledge management in BIM using ontology and semantic web technology. *Safety Science*, 87, 202-213.

29.Tserng, H. P., Yin, S. Y., Dzeng, R. J., Wou, B., Tsai, M. D., & Chen, W. Y. (2009). A study of ontology-based risk management framework of construction projects through project life cycle. *Automation in Construction*, 18(7), 994-1008.

30.Zhang, Sijie, Frank Boukamp, and Jochen Teizer. "Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA)." *Automation in Construction*, 52 (2015): 29-41.

31.Lee, Seul-Ki, Ka-Ram Kim, and Jung-Ho Yu. "BIM and ontology-based approach for building cost estimation." *Automation in Construction*, 41 (2014): 96-105.

32.Tang, P., Huber, D., Akinci, B., Lipman, R., & Lytle, A. (2010). Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in Construction*, *19*(7), 829-843.

33.Deng, Yichuan, Jack CP Cheng, and Chimay Anumba. "Mapping between BIM and 3D GIS in different levels of detail using schema mediation and instance comparison." *Automation in Construction,* 67 (2016): 1-21.

34.Irizarry, Javier, Ebrahim P. Karan, and Farzad Jalaei. "Integrating BIM and GIS to improve the visual monitoring of construction supply chain management." *Automation in Construction*, 31 (2013): 241-254.

35.Xu, Zhao, and Volker Coors. "Combining system dynamics model, GIS and 3D visualization in sustainability assessment of urban residential development." *Building and Environment,* 47 (2012): 272-287.

36.Patel, K. M., Patel, T. D., & Patel, P. J. (2017). Application of GIS in Construction Management. In International Conference on Construction, Real Estate, *Infrastructure and Project Management*, (pp. 104-112).

37.Kumar, A. Chaitanya, and T. Reshma. "4D applications of GIS in construction management." *Advances in Civil Engineering,* 2017 (2017).

38.Hardin, B., & McCool, D. (2015). *BIM and construction management: proven tools, methods, and workflows*. John Wiley & Sons.

39.Tang, L., Chen, C., Tang, S., Wu, Z., & Trofimova, P. (2017). Building Information Modeling and Building Performance Optimization. *Encyclopedia of Sustainable Technologies*.

40.Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., ... & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500-521.

41. (a)Sacks, R., Koskela, L., Dave, B. A., & Owen, R. (2010). Interaction of lean and building information modeling in construction. *Journal Of Construction Engineering and Management,* 136(9), 968-980.

42. (b) Sacks, R., Radosavljevic, M., & Barak, R. (2010). Requirements for building information modeling based lean production management systems for construction. *Automation in Construction*, 19(5), 641-655.

43.Korb, Samuel, and Rafael Sacks. "Towards multi-project simulation of a lean production system for customized apartment buildings." *2018 Winter Simulation Conference (WSC). IEEE*, 2018.

44.Zuo, Jian, and Zhen-Yu Zhao. "Green building research–current status and future agenda: A review." *Renewable and Sustainable Energy Reviews*, 30 (2014): 271-281.

45.Azhar, Salman, et al. "Building information modeling for sustainable design and LEED® rating analysis." *Automation in Construction,* 20.2 (2011): 217-224.

46.Wong, Johnny Kwok Wai, and Jason Zhou. "Enhancing environmental sustainability over building life cycles through green BIM: A review." *Automation in Construction*, 57 (2015): 156-165.

47.Jrade, A., and F. Jalaei. "Integrating building information modelling with sustainability to design building projects at the conceptual stage, *Building Simulation,* 6 (4): 429–444." (2013).

48. Grieves, M. (2014). Digital twin: manufacturing excellence through virtual factory replication. *White paper*, 1(2014), 1-7.

49.Hartmann, T., Van Meerveld, H., Vossebeld, N., & Adriaanse, A. (2012). Aligning building information model tools and construction management methods. *Automation in Construction*, *22*, 605-613.

50.Sacks, R., O. Rozenfeld, and Yehiel Rosenfeld. "Spatial and temporal exposure to safety hazards in construction." *Journal Of Construction Engineering and Management,* 135.8 (2009): 726-736.

51.Li, X., Yi, W., Chi, H. L., Wang, X., & Chan, A. P. (2018). A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in Construction*, *86*, 150-162.

52.Moon, D., Chung, S., Kwon, S., Seo, J., & Shin, J. (2019). Comparison and utilization of point cloud generated from photogrammetry and laser scanning: 3D world model for smart heavy equipment planning. *Automation in Construction*, *98,* 322-331.

53.Tuttas, S., Braun, A., Borrmann, A., & Stilla, U. (2016). Evaluation of acquisition strategies for image-based construction site monitoring. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *41*, 733-740.

54.Álvares, Juliana Sampaio, Dayana Bastos Costa, and Roseneia Rodrigues Santos de Melo. "Exploratory study of using unmanned aerial system imagery for construction site 3D mapping." *Construction Innovation,* (2018).

55.Zuo, Z., Cheng, J., Guo, H., & Li, Y. (2021). Knowledge mapping of research on strategic mineral resource security: A visual analysis using CiteSpace. *Resources Policy*, *74*, 102372.

56.Ozcan, Sercan, and Adnan Corum. "A scientometric analysis of remanufacturing by mapping scientific, organizational, and national concentration zones." *IEEE Transactions on Engineering Management,* 68.4 (2019): 1055-1071.

57.Eck, Nees Jan van, and Ludo Waltman. "How to normalize cooccurrence data? An analysis of some well-known similarity measures." *Journal Of The American Society for Information Science and Technology,* 60.8 (2009): 1635-1651.

58.Li, Y., Guo, J. E., Sun, S., Li, J., Wang, S., & Zhang, C. (2022). Air quality forecasting with artificial intelligence techniques: A scientometric and content analysis. *Environmental Modelling & Software*, 105329.

59.Xue, X., Yang, X., Deng, Z., Tu, H., Kong, D., Li, N., & Xu, F. (2021). Global Trends and Hotspots in Research on Rehabilitation Robots: A Bibliometric Analysis From 2010 to 2020. *Frontiers in Public Health*, *9*.

60.Wang, Tao, Azlan Shah Ali, and Cheong Peng Au-Yong. "Exploring a body of knowledge for promoting the building information model for facility management." *Ain Shams Engineering Journal,* 13.4 (2022): 101717.

61.Locatelli, M., Seghezzi, E., Pellegrini, L., Tagliabue, L. C., & Di Giuda, G. M. (2021). Exploring Natural Language Processing in Construction and Integration with Building Information Modeling: A Scientometric Analysis. *Buildings*, *11*(12), 583.

62.Sun, K., Liu, Y., Guo, Z., & Wang, C. (2016). Visualization for knowledge graph based on education data. *International Journal of Software and Informatics*, 10(3), 1-13.

63.Ozturk, Gozde Basak. "Trends in interoperability in building information modeling (BIM) research: A scientometric analysis of authors and articles.*" A/Z: ITU Journal of Faculty of Architecture* 17.3 (2020): 169-183.(a)

64.Ozturk, Gozde Basak. "Digital twin research in the AECO-FM industry." *Journal of Building Engineering,* 40 (2021): 102730.

65.Wu, Z., Chen, C., Cai, Y., Lu, C., Wang, H., & Yu, T. (2019). BIM-based visualization research in the construction industry: A network analysis. *International Journal of Environmental Research and Public Health*, *16*(18), 3473.

66.Ozturk, Gozde Basak. "Interoperability in building information modeling for AECO/FM industry.*" Automation in Construction,* 113 (2020): 103122.(b)

67.Baarimah, A. O., Alaloul, W. S., Liew, M. S., Kartika, W., Al-Sharafi, M. A., Musarat, M. A., ... & Qureshi, A. H. (2021). A bibliometric analysis and review of building information modelling for post-disaster reconstruction. *Sustainability*, *14*(1), 393.

68.Wang, Jian, and Fakhar Shahzad. "A Visualized and Scientometric Analysis of Health Literacy Research." *Frontiers in Public Health,* 9 (2021).

69.Baffoe, Gideon, and Keith Kintrea. "Towards Understanding the Landscapes of Neighbourhood Research: An Insight from Bibliometric Analysis." *Sustainability*, 14.5 (2022): 3115.

70.Wei, Fangwu, Tony H. Grubesic, and Bradley Wade Bishop. "Exploring the GIS knowledge domain using CiteSpace." *The Professional Geographer*, 67.3 (2015): 374-384.

71.Li, X., Wu, P., Shen, G. Q., Wang, X., & Teng, Y. (2017). Mapping the knowledge domains of Building Information Modeling (BIM): A bibliometric approach. *Automation in Construction*, *84*, 195-206.

72.Jin, R., Zou, Y., Gidado, K., Ashton, P., & Painting, N. (2019). Scientometric analysis of BIM-based research in construction engineering and management. *Engineering, Construction and Architectural Management*, *26*(8), 1750-1776.

73.Vilutiene, T., Kalibatiene, D., Hosseini, M. R., Pellicer, E., & Zavadskas, E. K. (2019). Building information modeling (BIM) for structural engineering: A bibliometric analysis of the literature. *Advances in Civil Engineering*, *2019*.

74.Chen, D., Xiang, P., Jia, F., & Guo, J. (2022). A systematic review of current status and trends of mega-infrastructure projects. Ain Shams Engineering Journal, 13(6), 101773.

75.Marzouk, M., & Elshaboury, N. (2022). Science mapping analysis of embodied energy in the construction industry. *Energy Reports,* 8, 1362-1376.

76.Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*, *12*(2), 1-257.

77.Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, *29*(12), 2724-2743.

78.Ringler, D., & Paulheim, H. (2017, September). *One knowledge graph to rule them all? Analyzing the differences between DBpedia, YAGO, Wikidata & co.* In Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz) (pp. 366-372). Springer, Cham.

79.Özçakir, Feridun Cemal, and A. Yılmaz ÇAMURCU. "Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması." *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6.12 (2007): 21-37.

80.Pegarkov, Daniel D. National Security Issues. *Nova Publishers*, 2006.

81.Bhardwaj, Brijesh Kumar, and Saurabh Pal. "Data Mining: A prediction for performance improvement using classification." arXiv preprint arXiv:1201.3418 (2012).

82.Bharati, Mrs, and M. Ramageri. "Data mining techniques and applications." (2010).

83.Berkhin, Pavel. "A survey of clustering data mining techniques." *Grouping multidimensional data*. Springer, Berlin, Heidelberg, 2006. 25-71.

84.Kameshwaran, K., and K. Malarvizhi. "Survey on clustering techniques in data mining." *International Journal of Computer Science and Information Technologies*, 5.2 (2014): 2272-2276.

85.Taniar, D. (Ed.). (2008). *Data mining and knowledge discovery technologies*. IGI Global.

86.Bharati, Mrs, and M. Ramageri. "Data mining techniques and applications." (2010).

87.Ahmad, Parvez, Saqib Qamar, and Syed Qasim Afser Rizvi. "Techniques of data mining in healthcare: a review." *International Journal of Computer Applications,* 120.15 (2015).

88.Gorade, Sudhir M., Ankit Deo, and Preetesh Purohit. "A study of some data mining classification techniques." *International Research Journal of Engineering and Technology*, 4.4 (2017): 3112-3115.

89.Gupta, Swati. "A multiple regression technique in data mining." *International Journal of Computer Applications,* 126.5 (2015).

90.Singh, Yashpal, and Alok Singh Chauhan. "NEURAL NETWORKS IN DATA MINING." Journal of Theoretical & Applied Information Technology 5.1 (2009).

91.Gaur, Priyanka. "Neural networks in data mining." *International Journal of Electronics and Computer Science Engineering*, (IJECSE, ISSN: 2277-1956) 1.03 (2012): 1449-1453.

92.Ni, Xianjun. "Research of data mining based on neural networks." *World Academy of Science, Engineering and Technology*, 39.1 (2008): 381-384.

93.Sharma, Meenakshi. "Data mining: A literature survey." *International Journal of Emerging Research in Management & Technology,* 3.2 (2014).

94.PK, FATHIMA ANJILA. "What is Artificial Intelligence?." "*Success is no accident. It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do*". (1984): 65.

95.Hong, Tzung-Pei, and Yeong-Chyi Lee. "An overview of mining fuzzy association rules." *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models* (2008): 397-410.

96.Al-Maolegi, M., & Arkok, B. (2014). An improved Apriori algorithm for association rules. *arXiv preprint arXiv:1403.3948.*

97.Jia, B., Dong, C., Chen, Z., Chang, K. C., Sullivan, N., & Chen, G. (2018, July). Pattern discovery and anomaly detection via knowledge graph. In *2018 21st International Conference on Information Fusion (FUSION)* (pp. 2392-2399). IEEE.

98.Castro, E. P., Maia, T. D., Pereira, M. R., Esmin, A. A., & Pereira, D. A. (2018). Review and comparison of Apriori algorithm implementations on Hadoop-MapReduce and Spark. *The Knowledge Engineering Review*, *33*.

99.Singh, Harpreet, Renu Dhir, and Munish Saini. "A Review of Various Methods for Association Rule Mining." *International Journal of Advanced Research in Computer Science,* 3.3 (2012).

100.Chee, C. H., Jaafar, J., Aziz, I. A., Hasan, M. H., & Yeoh, W. (2019). Algorithms for frequent itemset mining: a literature review. *Artificial Intelligence Review*, *52*(4), 2603-2621.

101.Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*(1), 1-37.

102.Zhao, Xianbo. "A scientometric review of global BIM research: Analysis and visualization." *Automation in Construction*, 80 (2017): 37-47.) (21)

103.Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc.

104.Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. *Nature,* 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2.

105.Reback, J., McKinney, W., Van Den Bossche, J., Augspurger, T., Cloud, P., Klein, A., ... & Seabold, S. (2020). pandas-dev/pandas: Pandas 1.0. 5. *Zenodo*.

106.John D. Hunter, "Matplotlib: A 2D Graphics Environment," in *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, May-June 2007, doi: 10.1109/MCSE.2007.55.

107.Toivonen, H. (2010). Apriori algorithm. In Encyclopedia of machine learning. Springer.

108.Hahsler M (2021). arulesViz: Visualizing Association Rules and Frequent Itemsets. R package version 1.5-1, https://CRAN.R-project.org/package=arulesViz.

109.Tan, Pang-Ning, Vipin Kumar, and Jaideep Srivastava. "Selecting the right objective measure for association analysis." *Information Systems*, 29.4 (2004): 293-313.

110.Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278–282).

111.Gerald, B. (2018). A brief review of independent, dependent and one sample t-test. *International Journal of Applied Mathematics and Theoretical Physics*, 4(2), 50-54.

112.Becerik-Gerber, B., Jazizadeh, F., Li, N., & Calis, G. (2012). Application areas and data requirements for BIM-enabled facilities management. *Journal Of Construction Engineering and Management*, 138(3), 431-442.

113.Sanhudo, L., Ramos, N. M., Martins, J. P., Almeida, R. M., Barreira, E., Simões, M. L., & Cardoso, V. (2020). A framework for in-situ geometric data acquisition using laser scanning for BIM modelling. *Journal of Building Engineering*, 28, 101073.

114.Amirebrahimi, S., Rajabifard, A., Mendis, P., & Ngo, T. (2016). A framework for a microscale flood damage assessment and visualization for a building using BIM–GIS integration. *International Journal of Digital Earth*, 9(4), 363-386.

115.Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction*, 114, 103179.

116.Cerovsek, T. (2011). A review and outlook for a 'Building Information Model'(BIM): A multi-standpoint framework for technological development. *Advanced Engineering Informatics*, 25(2), 224-244.

117.Webster, F. M. (1993). *The PMBOK® and PMP® exam*. PM Network, 7(12), 41–42.

118.Barlish, K., & Sullivan, K. (2012). How to measure the benefits of BIM—A case study approach. Automation in construction, 24, 149-159.

119.Zhang, S., Teizer, J., Lee, J. K., Eastman, C. M., & Venugopal, M. (2013). Building information modeling (BIM) and safety: Automatic safety checking of construction models and schedules. *Automation in Construction*, 29, 183-195.

120.Eadie, R., Browne, M., Odeyinka, H., McKeown, C., & McNiff, S. (2013). BIM implementation throughout the UK construction project lifecycle: An analysis. *Automation in Construction*, 36, 145-151.

121.Schlueter, A., & Thesseling, F. (2009). Building information model based energy/exergy performance assessment in early design stages. *Automation in Construction*, 18(2), 153-163.

122.Xiong, X., Adan, A., Akinci, B., & Huber, D. (2013). Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction*, 31, 325-337.

123.Thormark, C. (2002). A low energy building in a life cycle—its embodied energy, energy need for operation and recycling potential. Building and environment, 37(4), 429-435.

124.Gu, N., & London, K. (2010). Understanding and facilitating BIM adoption in the AEC industry. *Automation in Construction*, 19(8), 988-999.

125.Cheng, T., & Teizer, J. (2013). Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications. *Automation in Construction*, *34*, 3-15.

126.Guo, H., Yu, Y., & Skitmore, M. (2017). Visualization technology-based construction safety management: A review. *Automation in Construction*, *73*, 135-144.

T.C.

AYDIN ADNAN MENDERES UNIVERSITY

GRADUATE SCHOOLOF NATURAL AND APPLIED SCIENCES

SCIENTIFIC ETHICAL STATEMENT

I hereby declare that I composed all the information in my master's / doctoral thesis entitled within the framework of ethical behavior and academic rules, and that due references were provided and for all kinds of statements and information that do not belong to me in this study in accordance with the guide for writing the thesis. I declare that I accept all kinds of legal consequences when the opposite of what I have stated is revealed.

Gamze Gülsün GAZİ

… / 09/ 2022

CURRICULUM VITAE

Last name, First name: Gamze Gülsün GAZİ

Education

Level Institute Date of graduation	**Institute**	**Date of Graduation**
Bachelor’s Civil Engineering	Aydın Adnan Menderes University	2019