

mgr inż. Ewelina Mitera-Kielbasa^{1*)}
 ORCID: 0000-0002-1126-5798
 dr hab. inż. Krzysztof Zima, prof. PK¹⁾
 ORCID: 0000-0001-5563-5482

Exchange Information Requirements EIR in BIM projects: automatic classification based on ISO 19650 categories

Wymagania dotyczące wymiany informacji EIR w metodyce BIM – automatyczna klasyfikacja względem kategorii zgodnych z ISO 19650

DOI: 10.15199/33.2026.05.05

Abstract. The aim of this study was to develop a model for the automatic classification of fragments of Exchange Information Requirements (EIR) according to the categories defined in ISO 19650. TF-IDF, Word2Vec with SVM, and BERT were applied. Despite the limited dataset, the F1 score reached a value equal to or greater than 0.7. The study confirms the effectiveness of the approach and highlights the potential of AI in supporting the creation of customised EIRs for effective BIM implementation in construction projects.

Keywords: EIR; BIM; automation; AECO; AI.

Streszczenie. Artykuł dotyczy opracowania modelu automatycznej klasyfikacji fragmentów wymagań dotyczących wymiany informacji EIR zgodnie z kategoriami ISO 19650. Zastosowano TF-IDF, Word2Vec z SVM oraz BERT. Pomimo ograniczonego zbioru danych, uzyskano średnią wartość wskaźnika F1 równą lub wyższą niż 0,7. Badanie potwierdza skuteczność podejścia i wskazuje potencjał AI w usprawnieniu tworzenia spersonalizowanych EIR na potrzeby efektywnej implementacji BIM w przedsięwzięciach budowlanych.

Słowa kluczowe: EIR; BIM; automatyzacja; AECO; AI.

A 2023 study by the Royal Institution of Chartered Surveyors (RICS) indicates a stabilisation in the adoption of digitalisation within the built environment sector following a period of rapid growth during the COVID-19 pandemic, accompanied by an improvement in attitudes towards the implementation of Artificial Intelligence (AI). The main barriers remain cost, required labour input, organisational change, and a shortage of skilled professionals [1]. Building Information Modelling (BIM) is an example of digitalisation within the AECO sector (*Architecture, Engineering, Construction and Operation*), although challenges related to achieving full interoperability still remain [2]. While the adoption of BIM to date has delivered measurable benefits, its implementation still encounters resistance, particularly among small and medium-sized enterprises (SMEs) [3]. Interest in BIM continues to grow, including its incorporation into national regulations and public procurement frameworks – 35% of European countries already mandate or plan to mandate its use [4].

According to ISO 19650, a key component of information management within a construction project is the Exchange Information Requirements (EIR), which define what information should be delivered, when, in what format, and to which stakeholders [5]. The development of EIRs is a complex task requiring expert support; however, when properly developed, they can reduce inefficiencies and provide added

B adanie Royal Institution of Chartered Surveyors (RICS) z 2023 r. wskazuje na stabilizację we wdrażaniu cyfryzacji w sektorze środowiska zbudowanego po okresie dynamicznego wzrostu podczas pandemii COVID-19, przy jednoczesnej poprawie postaw wobec wdrażania sztucznej inteligencji (ang. *Artificial Intelligence* (AI)). Główne bariery to koszty, wymagany nakład pracy, zmiany organizacyjne i niedobór specjalistów [1]. Modelowanie informacji o obiekcie budowlanym (ang. *Building Information Modelling* (BIM)) jest przykładem wdrażania cyfryzacji w sektorze AECO (ang. *Architecture, Engineering, Construction and Operation*), choć nadal istnieją trudności w osiągnięciu pełnej interoperacyjności [2]. Dotychczasowa implementacja BIM przyniosła wymierne korzyści, lecz wdrażanie napotyka opór, przede wszystkim w sektorze MŚP [3]. Mimo tego zainteresowanie BIM zwiększa się, m.in. w przepisach krajowych i zamówieniach publicznych. Obecnie 35% państw europejskich już wymaga lub planuje jego stosowanie [4].

Zgodnie z ISO 19650, kluczowym elementem zarządzania informacją w procesie inwestycyjnym są Exchange Information Requirements (EIR), które określają, jakie informacje powinny zostać dostarczone, w jakim terminie, formie oraz którym interesariuszom [5]. Ich opracowanie jest złożone i wymaga wsparcia ekspertów, ale właściwie przygotowane mogą zapobiec nieefektywności i zapewnić wartość dodaną [6]. Mimo takiego znaczenia EIR, literatura przedmiotu nadal wykazuje niedobór spójnych, standaryzowanych i zgodnych z normami metod definiowania wymagań informacyjnych [7].

¹⁾ Politechnika Krakowska im. Tadeusza Kościuszki, Wydział Inżynierii Łądowej

^{*)} Correspondence address: ewelina.mitera@pk.edu.pl

value [6]. Despite their importance, the literature continues to lack consistent, standardised, and compliant methodologies for defining information requirements [7].

This paper presents a model for the automatic classification of EIR content according to the categories defined in ISO 19650. The proposed model forms part of a broader solution intended to support the development of customised EIRs.

Assumptions and methodology

According to the guidelines of ISO 19650, the Exchange Information Requirements (EIR) document specifies the information requirements imposed on project stakeholders in construction projects delivered using BIM. These requirements may be defined by the client’s team towards the design team, contractors, and principal contractors. The standard distinguishes three primary categories of such information:

- **managerial** – related to project management, including procedures for interdisciplinary coordination;
- **commercial** – relating to project objectives, including information delivery and exchange schedules (with the standard emphasising the need to avoid unnecessary data collection);
- **technical** – covering technical aspects such as file formats, software requirements, and the level of information need etc.

Exchange Information Requirements define the organisational and asset-related information requirements for a given construction project. The individual components of EIR are presented in Figure 1.

Due to the complexity of developing a comprehensive EIR, this study proposes the use of a supporting model that assists in its creation through the automatic generation of personalised proposals using Artificial Intelligence (AI) and a chatbot. A key component of this model is a **classification structure based**

W artykule zaprezentowano model automatycznej klasyfikacji fragmentów EIR względem kategorii zgodnych z ISO 19650, który stanowi część rozwiązania wspierającego tworzenie spersonalizowanych EIR.

Założenia i metoda badania

Zgodnie z wytycznymi ISO 19650, dokument Exchange Information Requirements zawiera specyfikację wymagań informacyjnych stawianych interesariuszom w przedsięwzięciach budowlanych realizowanych w BIM. Wymagania definiowane przez zespół inwestora mogą być skierowane do biura projektowego, wykonawcy robót budowlanych oraz generalnego wykonawcy. Norma wyróżnia trzy główne kategorie takich informacji:

- **zarządcze**, związane z zarządzaniem przedsięwzięciem, w tym procedurami koordynacji międzybranżowej;
- **biznesowe**, odnoszące się do celów czy np. harmonogramów dostarczenia i wymiany informacji (norma akcentuje potrzebę unikania gromadzenia zbędnych danych);
- **techniczne**, m.in. formaty plików, wymagania dotyczące oprogramowania oraz poziomy potrzeby informacyjnej.

Wymagania wymiany informacji precyzują wymagania informacyjne organizacji, danego przedsięwzięcia i aktywów. Poszczególne składowe EIR przedstawiono na rysunku 1.

Ze względu na złożoność opracowania dokumentu EIR, badanie opisywane w artykule zakłada wykorzystanie modelu wspomagającego jego tworzenie przez automatyczne generowanie spersonalizowanych propozycji z wykorzystaniem AI i chatbota. Kluczowym komponentem tego modelu jest **struktura klasyfikacyjna bazująca na trzech wymienionych kategoriach informacji**. Integracja klasyfikatora z generatorem spersonalizowanego tekstu oraz chatbotem będzie możliwa dzięki wykorzystaniu API, umożliwiającemu komunikację między systemami.

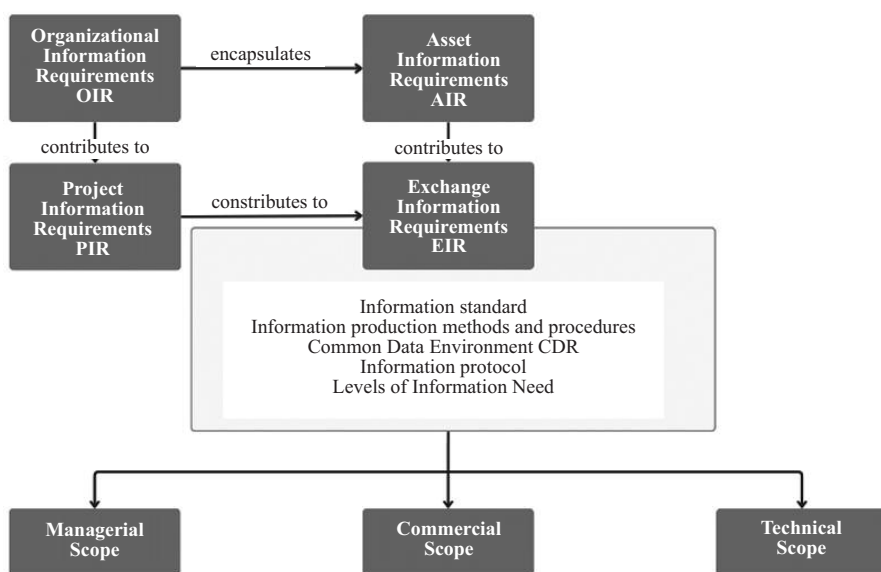


Fig. 1. Components of EIR
Rys. 1. Składowe EIR

Fig. elaboration based on ISO 19650-1 and ISO 19650-2
Rys. opracowano na podstawie ISO 19650-1 i ISO 19650-2

on the three aforementioned categories. The integration of the classifier with the personalised text generator and chatbot will be enabled through the use of an API, allowing communication between systems.

The study involved a series of experiments comparing the performance of various text classification methods, including: *Support Vector Machines* (SVM), the *Bidirectional Encoder Representations from Transformers* (BERT) model, the statistical analysis *Term Frequency–Inverse Document Frequency* (TF-IDF), and the Word2Vec model. The applied methods and models are illustrated in the text classification workflow in accordance with ISO 19650, as shown in Figure 2. The model was developed in Python using the Google Colab and JetBrains PyCharm environments. The input data consisted of EIR documents sourced from public and private procurement processes (Step 1). As a first step, the documents were pre-processed for automation by segmenting them into paragraphs (Step 2), followed by tokenisation – i.e. splitting into words – and normalisation (conversion to lowercase and removal of special characters and whitespace), using the NLTK library. For the BERT model, a dedicated tokenizer was applied (Step 3). In the next step text vectorisation was performed, converting textual data into numerical representations using TF-IDF (with the default *TfidfVectorizer* settings), Word2Vec (100-dimensional vectors, 5-word context window), and BERT (input sequences limited to 128 tokens). This stage resulted in the creation of a TF-IDF data matrix, Word2Vec feature vectors, and BERT vector embeddings (Step 4). The most critical stage of the process involved training the model to accurately assign text fragments to the appropriate categories. To this end, a subset of the available EIR documents was manually divided into paragraphs, each assigned a label corresponding to one of the categories K_A , K_B and K_C , representing managerial, commercial and technical information respectively. In the machine learning process, the model was trained using an 80:20 data split between training and validation datasets (Step 5). The SVM method was

W ramach badania przeprowadzono eksperymenty, porównujące skuteczność różnych metod klasyfikacji tekstu, obejmujące: metodę wektorów nośnych *Support Vector Machines* (SVM); dwukierunkowy model *Bidirectional Encoder Representations from Transformers* (BERT); analizę statystyczną *Term Frequency – Inverse Document Frequency* (TF-IDF) oraz model Word2Vec. Wykorzystywane metody zobrazowano na modelu procesu klasyfikacji tekstu względem ISO 19650 (rysunek 2). Model opracowano w języku Python z wykorzystaniem środowisk Google Colab oraz JetBrains PyCharm. Dane wejściowe stanowiły dokumenty EIR pochodzące z przetargów publicznych i prywatnych (krok 1). W pierwszej kolejności dokumenty poddano przygotowaniu do automatyzacji przez podział na akapity (krok 2), a następnie tokenizację, tj. podział na wyrazy i normalizację (małe litery, usunięcie specjalnych i białych znaków), wykorzystując bibliotekę NLTK. W przypadku modelu BERT zastosowano dedykowany tokenizator (krok 3). W kolejnym kroku przeprowadzono wektoryzację tekstu, przekształcając dane tekstowe w wartości liczbowe za pomocą TF-IDF (domyślna konfiguracja *TfidfVectorizer*), Word2Vec (100-wymiarowe wektory, kontekst 5 słów) i BERT (wejściowe sekwencje ograniczone do 128 tokenów). Wynikiem tego etapu były: macierz danych TF-IDF; wektory cech Word2Vec oraz reprezentacje BERT (krok 4). Najistotniejszym etapem procesu okazało się „nauczenie” modelu właściwej klasyfikacji fragmentów tekstu do odpowiednich kategorii. W tym celu część dostępnych dokumentów EIR podzielono ręcznie na akapity i przyporządkowano każdemu z nich odpowiednią etykietę (label), odpowiadającą jednej z kategorii K_A , K_B i K_C , czyli dotyczących aspektów zarządczych, biznesowych oraz technicznych. W procesie uczenia maszynowego model trenowano z wykorzystaniem danych podzielonych w stosunku 80 : 20 między przygotowanymi do treningu i pod walidację (krok 5). Wykorzystano metodę SVM (SVM z jądrem RBF, $c = 1$) i model BERT (trenowany przez 10 epok, batch size = 8). **Gotowe**

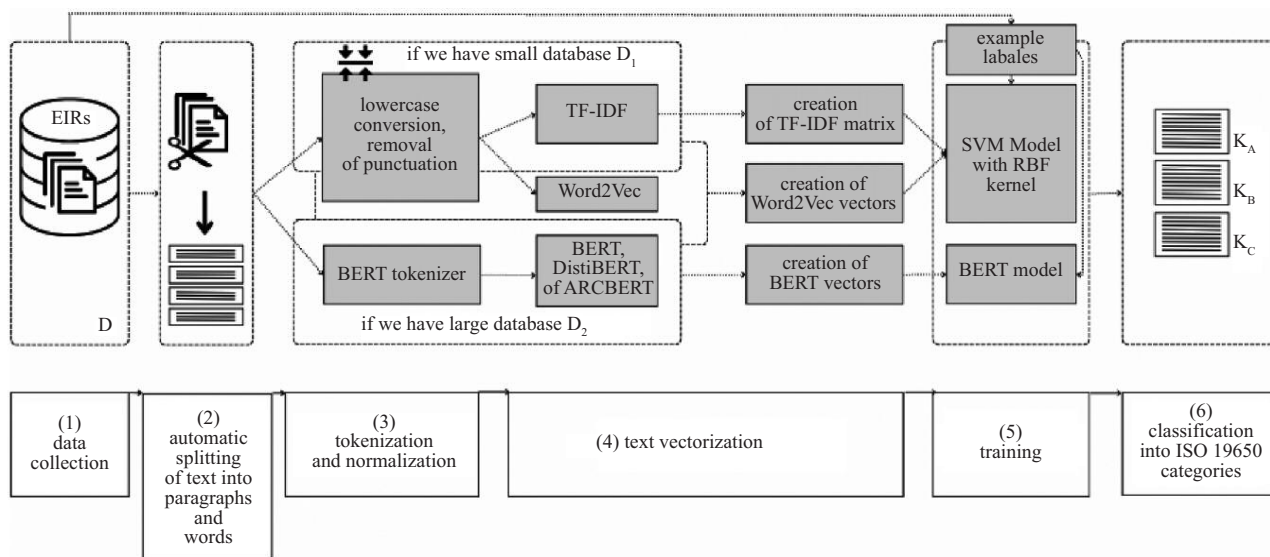


Fig. 2. Text classification process by ISO 19650 categories for EIR
Rys. 2. Proces klasyfikacji tekstu względem kategorii ISO 19650 w przypadku EIR

Fig. own study
Rys. opracowanie własne

applied (RBF kernel, $c = 1$), as well as the BERT model (trained for 10 epochs with a batch size of 8). **The resulting models were designed to automatically classify newly provided text fragments into the appropriate categories (Step 6).** The model was subsequently validated and its classification performance evaluated.

Results

To train the model, Exchange Information Requirements (EIR) were collected from both public and private construction projects. The data were pre-processed by segmenting the text into paragraphs ($n = 14,568$), each of which was assigned to one of the three information categories defined in ISO 19650: managerial (labelled as K_A), commercial (K_B) and technical (K_C). As illustrated in the workflow in Figure 2, the data underwent tokenisation, normalisation, and vectorisation. Subsequently, the dataset was split into training and validation subsets. The evaluation results of the models are presented in Figure 3.

The *F1 score*, which is the harmonic mean of precision (i.e. the proportion of correct positive predictions among all positive predictions) and recall (i.e. the proportion of actual positives correctly identified by the model), achieved average value of 0.7 or higher. Likewise, *accuracy* – the ratio of correct predictions to the total number of cases – also exceeded 0.7 across all evaluated models. Analysis of the classification results across K_A , K_B , and K_C demonstrated the highest performance for K_A , and the lowest for K_C (e.g. for Word2Vec + SVM: $F1 = 0.9$ for K_A , $F1 = 0.36$ for K_C). Additionally, category assignments (K_A , K_B or K_C) were verified for selected sentences. The results are presented in table. The majority of classifications were assigned to the managerial category, which corresponds with the distribution of the training data –

modele miały na celu automatyczną klasyfikację nowych fragmentów tekstu do odpowiednich kategorii (krok 6). Modele zostały poddane walidacji oraz ocenie skutecznej klasyfikacji.

Wyniki

Do trenowania modelu wykorzystano wymagania dotyczące wymiany informacji z publicznych i prywatnych przedsiębiorstw budowlanych. Dane przygotowano przez podział tekstów na akapity ($n = 14\,568$), z których każdy przypisano do jednej z trzech kategorii informacyjnych, określonych w normie ISO 19650: zarządczej (oznaczonej dalej jako K_A); biznesowej (K_B) i technicznej (K_C). Zgodnie ze schematem przedstawionym na rysunku 2, dane poddano tokenizacji, normalizacji i wektoryzacji. Następnie zestaw danych został podzielony na zbiór treningowy i walidacyjny. Wyniki ewaluacji modeli przedstawiono na rysunku 3.

Wskaźnik F1, będący średnią harmoniczną miary precyzji (ang. *precision*), tj. odsetka trafnych przewidywań spośród wszystkich pozytywnych klasyfikacji oraz czułości (ang. *recall*), czyli odsetka poprawnie rozpoznanych rzeczywistych przypadków, osiągnął we wszystkich przypadkach średnią wartość 0,7 lub większą. Z kolei dokładność (ang. *accuracy*), mierząca udział poprawnych klasyfikacji względem całkowitej liczby przykładów – przekroczyła wartość 0,7 w przypadku wszystkich testowanych modeli. Analiza wyników klasyfikacji w poszczególnych kategoriach K_A , K_B i K_C wykazała największą skuteczność kategorii K_A , a najgorszą K_C (np. w przypadku Word2Vec+SVM: $F1 = 0,9$ dla K_A oraz $F1 = 0,36$ dla K_C). Ponadto sprawdzono przypisanie kategorii K_A , K_B i K_C do wybranych zdań. Wyniki przedstawiono w tabeli. Większość przyporządkowań dotyczyła **kategori informacji zarządczych**, co jest spójne z rozkładem danych. Ta katego-

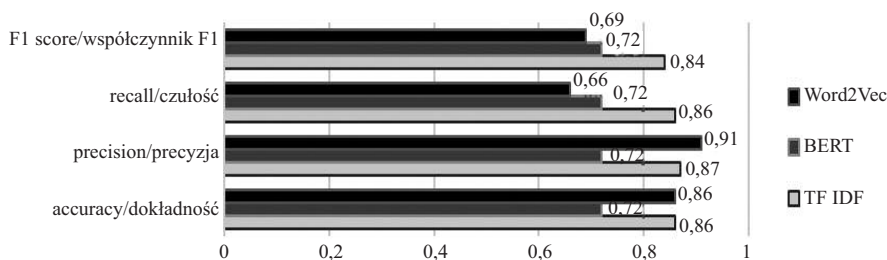


Fig. 3. Evaluation metrics of the text classification model for EIR
Rys. 3. Metryki ewaluacji modelu klasyfikacji tekstu w przypadku EIR

Fig. own study
Rys. opracowanie własne

Model’s predicted categories for provided EIR sentences

Przewidywane kategorie modelu w przypadku wybranych zdań EIR

Own study
Opracowanie własne

Sentence/Zdanie	TF-IDF + SVM	Word2Vec + SVM	BERT
The contractor shall compile in the BEP, for approval, the file formats that will be delivered to the client at each stage of the project/Wykonawca powinien zestawić w Planie Wykonania BIM (BEP), do zatwierdzenia, formaty plików, które będą przekazywane Zamawiającemu na każdym etapie projektu)	K_A	K_A	K_C
The final BEP shall include information related to 'Project Roles & Responsibilities/Końcowy Plan Wykonania BIM (BEP) powinien zawierać informacje dotyczące „Ról i odpowiedzialności”)	K_A	K_A	K_A
All issues should be promptly reported to the client in the common data environment/Wszystkie problemy i niezgodności powinny być niezwłocznie zgłaszane Zamawiającemu we wspólnym środowisku danych (CDE)	K_A	K_A	K_C

this category contained the largest number of entries (8,494 paragraphs). This reflects industry practice, as managerial requirements typically represent the most extensive component of information requirements placed upon project stakeholders.

This study presented in this paper investigates the application of Artificial Intelligence (AI) tools and methods to support the delivery of construction projects developed in accordance with the BIM methodology. The prepared dataset was subjected to text vectorisation using TF-IDF, Word2Vec and BERT. The first method is based on statistical analysis, evaluating the frequency of a word within a document relative to its frequency across the entire corpus. Its main advantage lies in its effectiveness even when working with relatively small datasets, which was one of the reasons for its inclusion in this study. Word2Vec also performs well in contexts involving limited data and enables the semantic meaning of words to be considered within their contextual proximity. Unlike BERT, which analyses word meaning in relation to the entire sentence, Word2Vec primarily considers neighbouring words and their local semantic context. The Bidirectional Encoder Representations from Transformers (BERT) model, developed by Google, utilises a bidirectional transformer architecture. This advanced tool handles both the classification and vectorisation of text and tends to achieve the best results when trained on large and diverse datasets.

In this study, the prepared texts were manually annotated with labels corresponding to information categories (K_A , K_B , K_C), in line with the ISO 19650 standard. These annotated data were used to train the models, which were then subjected to validation and performance evaluation. For the SVM method, which was also employed for classification, a radial basis function (RBF) kernel was used to effectively address non-linearly separable classification problems.

The classification results across all three approaches were promising, even though the dataset was relatively small compared to those typically used in AI-related research. Example category assignments for selected sentences are presented in table. In the first example, the word “format” might suggest a technical classification, as it refers to required file formats. However, the full sentence context indicates that it more appropriately belongs to the managerial category, as it concerns planning and approval procedures within the BIM Execution Plan (BEP). Such cases highlight the challenges of contextual interpretation by language models, particularly when trained on limited data. The predominance of classifications into the managerial category (K_A) reflects the distribution of the training data – this category contained the highest number of annotated paragraphs (8,494), which aligns with typical EIR content, where managerial requirements are the most extensively documented. As a result, the model had more opportunity to learn characteristic features of this category. The lowest accuracy was observed for predictions in the technical category (K_C), which is consistent with its lower representation in the training data. These observations

ria stanowiła największy zbiór w analizowanej bazie (8 494 akapity). Jest to zgodne z praktyką, ponieważ wymagania zarządcze obejmują zwykle najszerszą część wymagań informacyjnych nakładanych na interesariuszy.

Opisane w artykule badanie dotyczy zastosowania narzędzi i metod sztucznej inteligencji (AI) w celu usprawnienia realizacji przedsięwzięć budowlanych prowadzonych zgodnie z metodyką BIM. Przygotowaną wcześniej bazę danych poddano wektoryzacji tekstu z wykorzystaniem TF-IDF, Word2Vec i BERT. Pierwsza z nich polega na analizie statystycznej, oceniając częstość występowania danego słowa w dokumencie względem jego częstości w całej bazie danych. Główną zaletą tej metody jest jej skuteczność nawet przy niewielkim zbiorze danych, co stanowiło jeden z powodów jej zastosowania w badaniu. Word2Vec dobrze radzi sobie również w sytuacjach, gdy baza danych nie jest obszerna, a dodatkowo pozwala na uwzględnienie znaczenia słów w ich semantycznym kontekście. W przeciwieństwie do BERT, który analizuje znaczenie słowa w odniesieniu do całego zdania, Word2Vec bierze pod uwagę głównie sąsiednie wyrazy. Model BERT, będący jedną z nowoczesnych implementacji opracowanych przez Google, wykorzystuje architekturę transformatorów dwukierunkowych. Jest to zaawansowane narzędzie, które oprócz samej klasyfikacji, odpowiada również za wektoryzację tekstu. W praktyce BERT osiąga najlepsze wyniki, gdy dysponuje dużym i zróżnicowanym zbiorem danych treningowych.

W ramach badania przygotowane teksty zostały ręcznie oznaczone kategoriami informacyjnymi K_A , K_B , K_C , zgodnymi z normą ISO 19650. Dane te posłużyły do treningu modeli, które następnie poddano procesom walidacji i oceny skuteczności klasyfikacji. W przypadku metody SVM, którą również wykorzystano do klasyfikacji, zastosowano jądro funkcji radialnej (RBF) pozwalające skutecznie rozwiązywać problemy nieliniowo-separowalne.

Wyniki klasyfikacji w przypadku wszystkich trzech podejść okazały się obiecujące, nawet pomimo ograniczonej wielkości bazy danych – mniejszej niż w typowych badaniach z zakresu sztucznej inteligencji. Przykładowe wyniki przypisania kategorii do wybranych zdań przedstawiono w tabeli. W pierwszym przykładzie słowo „format” mogło sugerować przyporządkowanie do kategorii technicznej, ponieważ odnosi się do oczekiwanych formatów plików. Niemniej jednak kontekst całego zdania wskazuje raczej na aspekt zarządczy, ponieważ dotyczy planowania i zatwierdzania działań w ramach planu wykonania BIM – ang. *BIM Execution Plan* (BEP). Tego typu przypadki pokazują wyzwania w interpretacji kontekstu przez modele językowe, szczególnie przy ograniczonej liczbie danych. Dominacja klasyfikacji do kategorii zarządczej (K_A) może wynikać z tego, że przypisane do niej zdania były najliczniejsze w zbiorze treningowym (8 494 akapity), zwykle zresztą stanowi ona najobszerniejszą część dokumentów EIR. Model miał więc większe możliwości „nauczenia się” cech charakterystycznych tej kategorii. Najmniejszą dokładność zaobserwowano w przypadku przypisania do kategorii technicznej (K_C), co jest zgodne z faktem, że była ona najmniej reprezentowana w zbiorze treningowym. Te

reinforce the importance of a balanced dataset in training classification models. In EIR documents, the technical section is often less extensive than the other categories, and its most information-rich component, the Level of Information Need section, is typically provided as a separate appendix to the EIR document. Although the analysis of only three individual sentences (Table 1) does not constitute a representative sample, it does offer illustrative insights (which should be considered alongside the metrics in Figure 2). It indicates that BERT, despite its advanced architecture, performed less effectively than the simpler model, most likely due to the insufficient volume of training data. Nonetheless, given its potential, this model will be revisited in future studies using a more comprehensive and diverse corpus. In future research, it would be worthwhile to analyse classification performance based on smaller textual units. Shorter fragments, such as individual sentences, may yield higher F1 scores, particularly for models like Word2Vec+SVM or BERT, which perform better with semantically clear structures. However, a holistic approach is necessary, taking into account the potential loss of broader semantic context, which is especially relevant in further stages of personalised EIR generation and their automatically generated structure.

Conclusions

The ongoing digitalisation of the construction sector motivates the development of solutions that support its effective implementation. One of the key components of information management in BIM is the Exchange Information Requirements (EIR), namely the information exchange requirements imposed on stakeholders involved in a construction project. Although well-defined EIRs significantly reduce the risk of misunderstandings and support the achievement of project objectives, their development is a complex process – covering the entire life cycle of the built asset and encompassing managerial, commercial, and technical aspects. There remains a shortage of tools and templates aligned with current standards that would support the identification, classification, and structuring of information requirements. Moreover, limited expertise on the part of clients often results in ineffectively formulated EIRs, which may negatively affect overall project delivery.

In response to the identified challenges, the authors proposed an approach based on the use of Artificial Intelligence (AI) methods and tools to support the development of Exchange Information Requirements (EIR). The developed model is part of a broader automated system aimed at generating project-specific, customised EIRs tailored to the needs of the project and its stakeholders. **This study focused on a key stage of that process: the automatic classification of text fragments according to the information categories defined in ISO 19650 – namely managerial, commercial, and technical.** This classification provides the basis for further content structuring and generating of EIR documents with appropriate structure and informational value.

obserwacje potwierdzają znaczenie zrównoważonego rozkładu danych w procesie uczenia modeli klasyfikacyjnych. W EIR kategoria ta zwykle jest także mniej obszerna w porównaniu z pozostałymi, a bogata w informacje część dotycząca poziomów potrzeby informacyjnej stanowi zwykle odrębny załącznik do dokumentu. Analiza trzech pojedynczych zdań (tabela) nie stanowi wprawdzie reprezentatywnej próby, ale obrazuje, że BERT, mimo swojego zaawansowania, poradził sobie relatywnie gorzej niż prostsze modele, prawdopodobnie z powodu niedostatecznego rozmiaru danych treningowych. Ze względu na potencjał, model ten zostanie ponownie przebadany w przyszłości z wykorzystaniem bardziej rozbudowanej bazy danych. W dalszych badaniach warto również przeanalizować skuteczność klasyfikacji przy podziale danych na mniejsze jednostki tekstowe. Krótsze fragmenty, takie jak pojedyncze zdania, mogą prowadzić do uzyskania większych wartości wskaźnika F1, szczególnie w przypadku modeli, takich jak Word2Vec+SVM czy BERT, które dobrze sobie radzą z jednoznacznymi semantycznie strukturami. Należy jednak podejść do tego holistycznie, uwzględniając możliwość utraty szerszego kontekstu semantycznego, szczególnie istotnego w dalszych etapach badań nad generowaniem spersonalizowanych dokumentów EIR oraz ich automatycznie tworzoną strukturą.

Wnioski

Postępująca cyfryzacja sektora budowlanego motywuje do opracowania rozwiązań wspierających jej skuteczne wdrażanie. Jednym z kluczowych elementów zarządzania informacją w środowisku BIM są Exchange Information Requirements (EIR), czyli wymagania wymiany informacji wobec interesariuszy przedsięwzięcia budowlanego. Prawidłowo zdefiniowane EIR znacznie ograniczają ryzyko nieporozumień i umożliwiają lepsze osiągnięcie celów inwestycyjnych, ale ich opracowanie jest procesem złożonym, ponieważ obejmuje cały cykl życia obiektu oraz wiele aspektów zarządczych, biznesowych i technicznych. Na rynku nadal brakuje narzędzi i szablonów zgodnych z obowiązującymi normami, które wspierałyby identyfikację, klasyfikację i strukturyzację wymagań informacyjnych. Dodatkowo, ograniczenia kompetencyjne po stronie zamawiających prowadzą często do nieefektywnego formułowania EIR, co negatywnie wpływa na przebieg całego przedsięwzięcia.

W odpowiedzi na zidentyfikowane wyzwania zaproponowaliśmy podejście polegające na wykorzystaniu metod i narzędzi sztucznej inteligencji (AI) w celu wsparcia procesu opracowywania dokumentów Exchange Information Requirements (EIR). **Opracowany model stanowi część zautomatyzowanego systemu wspomagającego tworzenie spersonalizowanych EIR, dostosowanych do potrzeb konkretnego przedsięwzięcia i interesariuszy.** W ramach przeprowadzonego badania skoncentrowano się na kluczowym etapie tego procesu, jakim jest automatyczna klasyfikacja fragmentów tekstu wg kategorii informacyjnych określonych w normie ISO 19650, tj. zarządczej, biznesowej oraz technicznej. Klasyfikacja ta stanowi podstawę do dalszego uporządkowania

The study explored various approaches to text vectorisation and classification, including: TF-IDF, Word2Vec combined with SVM, and the BERT model. Despite the limited size of the dataset, all methods achieved satisfactory results, with the average F1 scores reached or exceeded 0.7. Particularly promising were the Word2Vec + SVM combination and the implementation of BERT, due to their capacity for capturing semantic context.

The findings highlight specific challenges associated with the digital transformation of the AECO sector. Effective information management is becoming a prerequisite for unlocking the full potential of BIM. Properly prepared EIRs can significantly enhance the efficiency, quality, and transparency of project delivery processes.

Received: 07.01.2026

Revised: 23.03.2026

Published: 21.05.2026

treści i wygenerowania dokumentu EIR o właściwej strukturze i wartości informacyjnej.

W ramach badania zastosowano różne podejście do wektoryzacji tekstu i klasyfikacji, w tym TF-IDF, Word2Vec z SVM oraz model BERT. Pomimo ograniczonego rozmiaru bazy danych, wszystkie metody osiągnęły zadowalające wyniki, a średnia wartość współczynnika F1 wyniosła lub przekroczyła 0,7. Szczególnie obiecujące wydaje się połączenie Word2Vec z SVM oraz implementacja modelu BERT ze względu na ich zdolność do uwzględniania kontekstu semantycznego.

Wyniki badania wskazują na konkretne wyzwania związane z cyfryzacją sektora AECO. Skuteczne zarządzanie informacją staje się warunkiem koniecznym pełnego wykorzystania potencjału BIM. Właściwe przygotowanie dokumentu EIR może realnie przełożyć się na zwiększenie efektywności, jakości i transparentności procesów inwestycyjnych.

Artykuł wpłynął do redakcji: 07.01.2026 r.

Otrzymano poprawiony o recenzjach: 23.03.2026 r.

Opublikowano: 21.05.2026 r.

Literature

[1] Sawhney A, Knight A. Digitalisation in construction report 2024. London, UK, Nov. 2024.

[2] Borkowski AS. Propedeutyka BIM – filozofia modelowania informacji o obiekcie budowlanym, 1st ed. Warsaw, Poland: Oficyna Wydawnicza Politechniki Warszawskiej, 2024.

[3] Awwad KA, Shibani A, Ghostin M. Exploring the critical success factors influencing BIM level 2 implementation in the UK construction industry: the case of SMEs. *International Journal of Construction Management*. 2022. DOI: 10.1080/15623599.2020.1744213.

[5] Mitera-Kielbasa E, Zima K. BIM Policy Trends in Europe: Insights from a Multi-Stage Analysis. *Applied Sciences*. 2024. DOI: 10.3390/app14114363.

[5] International Organization for Standardization, ISO 19650-1: 2018 Organization and digitization of information about buildings and civil engineering works, including BIM – Information management using building information modelling – Part 1: Concepts and principles. Switzerland, Geneva, 2018.

[6] Olawumi TO, Chan DWM. Building information modelling and project information management framework for construction projects. *Journal of civil engineering and management*. 2019. DOI: 10.3846/jcem.2019.7841.

[7] Oliveira A, Granja J, Bolpagni M, Motamedi A, Azenha M. Development of standard-based information requirements for the facility management of a canteen. *Journal of Information Technology in Construction*. 2024. DOI: 10.36680/j.itcon.2024.014.