ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USING SENTENCE TRANSFORMERS FOR SELF-ASSESSMENT IN DIGITAL TRANSFORMATION

BACHIRA ABOU EL KARAM[1], TARIK FISSAA[1] , RABIA MARGHOUBI[1]

[1] Innovative Research on Software Systems and data, INPT, Rabat, Morocco

E-mail: [1] abouelkaram.bachira@doctorant.inpt.ac.ma, [1] fissaa@ inpt.ac.ma, [1] m.rabia@ inpt.ac.ma

ABSTRACT

Digital transformation requires rethinking a company's organization to identify the necessary changes for implementing digital initiatives. It goes beyond technology, encompassing corporate strategy and impacting organizational culture, employee involvement, customer orientation, and business models. To embark on a digital transformation project, companies must first assess their current state regarding strategy, digital maturity, and organizational culture. Existing evaluation methods rely either on consulting services, which are effective but costly for SMEs, or closed-response questionnaires, which are quicker and standardized but limit expression, potentially introducing biases. To address these challenges and find a compromise between the affordability and the precision, this paper proposes an automated self-assessment approach based on open-ended responses, leveraging Sentence Transformers to evaluate and score SMEs' current state. Since the aforementioned evaluation requires high precision due to strategic decisions and investments that are resulting, and characterized by the diversity of unrestricted responses, particularly in a francophone context where cultural and linguistic nuances can significantly influence results, the approach must be tested and compared to the manual method which is often considered as a reference. To achieve this, a case study was conducted on a Moroccan SME that had previously been audited manually by a consulting firm, and the open-ended responses from this audit were subsequently analyzed automatically using a Sentence Transformer-based method as well as using statistical techniques: TF-IDF and LSA. The results demonstrated a strong alignment between the proposed approach and expert evaluations, confirming its effectiveness as a cost-efficient and scalable solution for SMEs, while outperforming other evaluated methods.

Keywords: *Digital Transformation, Self-Assessment, Neural network transformers, NLP, Semantic Similarity, Decision Making.*

1. INTRODUCTION

In a context marked by a constantly changing economic environment with the advent of Industry 4.0, increased competition and rapid technological advances, the transition to digital transformation is no longer simply an asset but an absolute necessity for companies to maintain their competitiveness and guarantee a robust online presence. [1] defines digital transformation as a systemic change in internal processes, business models and skills, using technology to deliver intelligent products and services that meet customer expectations. According to this definition, digital transformation goes beyond the digitization of resources and involves the transformation of structural and organizational aspects [2]. This explains why many companies that invest in and focus on individual technologies to solve momentary problems fail to generate the expected operational performance. Digital Transformation (DT) is characterized as multidisciplinary [3] and a holistic form of transformation enabled by information systems that are accompanied by fundamental economic and technological changes at both organizational and industry levels [4] [5] [6].

Therefore, to achieve a successful digital transformation project, companies must first analyze and assess in detail every aspect of the business, including its processes, technologies, human resources and organizational culture, to determine their requirements for implementing digital initiatives [7]. This evaluation must be conducted rigorously during the initial phase of the digital transformation project, referred to as the company's current state assessment. An accurate assessment is critical, as the results dictate the roadmap for digital initiatives. An erroneous evaluation can lead to incompatible digital projects and result in wasted investments. As highlighted by several studies [8] [9] [10] [11] [12], this evaluation must be multidimensional, encompassing a strategic audit to analyze the company's strategy and objectives, a digital maturity audit to assess the extent of technological integration, and a cultural audit to evaluate the organization's readiness for digital transformation and its capacity to adapt to change.

Previous studies on current state assessment for SMEs have primarily relied on expert-driven audits through interviews with stakeholders [12] [13][14], or by using self-assessment tools with closed-ended questionnaires [9][16][17][18]. While interviews provide valuable and detailed insights, they are often prohibitively expensive for SMEs due to financial constraints, rendering this option inaccessible for many. As a more cost-effective alternative, self-assessment tools with closed-ended questions have become increasingly popular. These tools offer a concise view of the company's strategic orientations and are easy to evaluate using reference answer scores. However, they constrain respondents' ability to express themselves fully, which can lead to biased or incomplete assessments.

These limitations highlight the need for a more flexible and cost-effective approach that allows companies, especially SMEs, to evaluate their digital readiness without financial burden or restrictive response formats.

To address these issues, this study seeks to answer the following research questions:

- How can open-ended responses improve the accuracy and richness of digital maturity assessments for SMEs?
- What are the specific challenges faced by SMEs in existing assessment methods, and how can a new approach address these challenges?
- In what ways can AI-driven semantic analysis enhance the evaluation process of digital readiness in organizations?

Building on these research questions, this study proposes a self-assessment approach based on open-ended responses, specifically designed for SMEs. Unlike traditional methods that rely on closed-ended questionnaires or expert evaluations, this approach allows companies to freely express their insights while automating the assessment process.

The main feature of this method lies in the use of Sentence Transformers, a state-of-the-art deep learning model in natural language processing. Instead of relying on experts to interpret and score responses, this model automatically analyzes user inputs and measures their semantic similarity to predefined reference answers. By combining the depth and flexibility of open-ended responses with the objectivity and consistency of AI-driven evaluation, this approach provides SMEs with a more accessible, scalable, and reliable solution to assess their digital maturity.

While the proposed methodology is generalized to various types of audits, the case study presented in this article focuses specifically on the strategic audit. For SMEs, this system will be a tool for evaluating the results of audit surveys, saving time, effort and money without diminishing quality, and can subsequently be integrated into the overall system for automating the digital transformation process.

The document is structured into six main sections. The second section provides a review of previous work relevant to this research. The third section offers an overview of Semantic Similarity in Natural Language Processing (NLP). The fourth section explains the methodology adopted in detail, followed by the fifth section, which presents proof of concept for the proposed approach. Lastly, the sixth section summarizes the conclusions drawn from the study and highlights potential future research directions.

2. RELATED WORK

This section begins by reviewing research articles that concentrate on tools and methods for assessing a company's current state in the context of digital transformation, a selection crucial for establishing a foundational understanding of existing assessment methods. Additionally, it explores articles that focus on the analysis and automated evaluation of open-ended responses in other sectors. This dual focus is essential for providing a comprehensive understanding of current evaluation approaches and for investigating how open-ended responses can enhance the evaluation process.

Several studies emphasize the role of consultancy services in evaluating a company's current state in the context of digital transformation, particularly for SMEs that lack internal expertise. Jukka and Pasi [10] introduce three tools—DIGI MATURITY, DIGI SWOT, and DIGI TRIANGLE—that rely on questionnaires analyzed by external organizations, complemented by interviews and facilitated workshops with company representatives. Similarly, Stich and Zeller [12] utilize the Acatech Industry 4.0 Maturity Index, which combines questionnaires with interviews conducted by competence centers to evaluate

SMEs' strategies and environments. Cimini and Pinto [13] propose the Digital Readiness Level 4.0 (DRL 4.0) model, incorporating 46 closed-ended questions, while enriching the process with semi-structured interviews and expert observations to assess digital readiness and develop tailored Industry 4.0 adoption roadmaps. Soluk and Nadine [14] investigate digital transformation in family businesses through semi-structured interviews and triangulation of data from interviews, observations, and additional sources. Ulas [15] highlights the reliance of Turkish SMEs on consulting services from government agencies or research institutes to effectively implement data-driven digital transformation strategies.

Although these consultant-led approaches offer tailored insights that address specific business needs, they are resource-intensive in terms of time and cost, making them less accessible to SMEs. To overcome the limitations of consultant-led assessments, many researchers have developed self-assessment tools, allowing organizations to independently evaluate their digital maturity. Olli and Jukka [9] propose a digital maturity framework and self-assessment tool focused on six dimensions: Strategy, Business Model, Customer Interface, Organization and Processes, People and Culture, and Information Technology. Each dimension of the maturity model comprises questions with response options that measure an organization's level of digitization. Santos and Martinho [16] introduced a holistic digital maturity assessment model, implemented through a self-administered questionnaire with closed-ended questions. Similarly, Haryanti and Rakhmawati [17] suggest an extended model, the DX-Self Assessment Maturity Model Tool, which evaluates seven dimensions—Organizational Structure, Technology, Strategy, Employees, Customers, Transformation Process, and Culture—using 4 Likert Scale closed-ended questions covering the 7 dimensions mentioned above. Leila Saari, Olli Kuusisto, and Juha Häikiö [18] present the ManuMaturity tool, designed specifically for self-assessment to help manufacturing companies progress in their digitalization and achieve Industry 4.0 goals. This tool employs a maturity model with seven dimensions, each containing themes and predefined questions that correspond to different maturity levels. However, a major limitation of these self-assessment tools is their reliance on closed-ended questions, which may not fully capture the nuances of an organization's digital maturity. Respondents are often constrained by predefined answer choices, potentially leading to biased or incomplete assessments. Furthermore, these tools lack adaptability to the diverse strategic contexts of SMEs, particularly in francophone environments.

Additionally, as highlighted in [9], SMEs often face difficulties in using these assessment tools due to the rigid nature of closed-ended responses, which limit their ability to express nuanced insights about their digital maturity. The predefined answer choices may not always align with the specific context of each company, leading to assessments that fail to capture their true level of digital readiness. Moreover, the terminology used, the formulation of questions, and the lack of contextualized examples can further hinder the accuracy of responses, especially for SMEs with varying degrees of digital literacy. Such constraints can result in misinterpretations and an oversimplification of complex digital transformation challenges.

To address these gaps, our approach introduces an automated, AI-driven evaluation method based on the analysis of open-ended responses. This allows respondents to express their digital transformation challenges more freely, while ensuring a robust and scalable assessment process through Natural Language Processing (NLP).

This approach enables a precise identification of the company's strategic priorities through a strategic audit, as well as a rigorous evaluation of its true digital maturity level through a digital maturity audit. These insights serve as a foundation for developing a clear roadmap that aligns the company's priority objectives with the most suitable digital technologies to achieve them.

The concept of automatic evaluation of open-ended responses has been explored in various fields, particularly in education, where it has been used to assess students' work. The approaches adopted range from traditional statistical models, such as Latent Semantic Analysis (LSA) and TF-IDF, to more advanced methods relying on neural networks incorporating the attention mechanism, known as Transformers.

For instance, in [19], the authors developed an online learning interface for the automatic evaluation of students' writing tests in Indonesia using Latent Semantic Analysis (LSA). This system achieved an accuracy of 83.3% compared to manual evaluations performed by teachers. In [20], the automatic evaluation of open-ended responses was carried out using a model combining specific and general information, coupled with an LSTM neural network to effectively capture word sequences. This approach demonstrated superior performance compared to existing models such as Logistic Regression (LR), Naïve Bayes (NB), and Decision Trees (DT).

More recently, many studies on the automatic evaluation of open-ended responses have turned to deep learning methods using state-of-the-art Transformer models. For example, the authors in [21] explored the automatic evaluation of short answers using various vector representation techniques, including Sentence-BERT (SBERT). The authors demonstrated that SBERT outperformed traditional approaches such as Word2Vec and Bag-of-Words in modeling student responses. The authors in [22] compared three variants of sentence-transformer models for the automatic grading of students' responses in a secondary school online learning system. Their study focused on the similarity between the model-generated grades and those assigned by teachers while considering processing time. Authors in [23] proposed an automatic essay grading system based on fine-tuning the BERT model, adopting a multi-task learning (MTL) approach to assess texts across multiple dimensions.

Although these studies have demonstrated the effectiveness of Transformers in scoring open-ended responses across various fields, particularly in education, they have not addressed critical contexts such as digital transformation. Specifically, the assessment of an organization's current state requires greater rigor and precision to guide strategic decisions on projects and investments, as well as the capacity to handle intrinsically diverse and unbounded responses. Furthermore, these studies have not been tested on French-language datasets, limiting their applicability in francophone contexts. This is the perspective from which our contribution emerges, aiming to propose a self-assessment approach for evaluating the current state of organizations. This approach leverages a multilingual variant of Sentence Transformers, to overcome the limitations of manual methods and those based on closed-ended responses previously discussed. We also evaluate the validity of this approach in a critical and francophone context by comparing it to the reference (manual) method and other statistical approaches.

3. BACKGROUND

This section presents an overview of the approaches employed to measure semantic similarity in natural language processing.

3.1 Statistical Models-Based Approaches

Statistical model-based approaches to measuring textual similarity work by first transforming the text into numerical vectors and then calculating the similarity between these vectors. There are several text vectorization techniques for transforming text data into numerical representations [24] [25] [26]: Bag-of-Words (BOW) is a technique that creates a set of unique words from the text corpus. Each document is represented by a vector where each dimension corresponds to a word, and the value indicates the frequency of occurrence of the word in the document. TF-IDF (Term Frequency-Inverse Document Frequency) is an enhancement to BOW. It considers the frequency with which a term appears in a document (TF), as well as its overall importance in all documents (IDF). N-grams: n-grams are sequences of n consecutive words in a text. N-grams can be used as features to represent documents. Bigrams (n=2) and trigrams (n=3) are the most used.

Once the vectors have been obtained, the similarity between two vectors (and therefore between the texts they represent) can be calculated using several formulas such as cosine similarity, Euclidean distance, Jaccard distance and many others... However, these techniques do not consider the sequence of words, and fail to capture the semantic relationships between words, as they generally rely on the number of occurrences of the corresponding word in the text This means that two sentences using different words but with a similar meaning may not be recognized as close in semantic space.

To overcome this limitation, Susan T. Dumais and George W. Furnas [27], introduced Latent Semantic Analysis (LSA), a technique for deducing the sense of words from the contexts in which they appear within large sets of texts. The fundamental principle of LSA is based on the idea that words that appear frequently in similar contexts tend to share similar meanings [28] [29]. Unlike traditional models of measuring semantic similarity based on the bag-of-words (BoW) technique, where the similarity between two documents is reduced to zero if they share no common terms, LSA goes beyond the words themselves to understand the underlying concepts. It works as follows (see figure

1) [26] [28] [29] [30] [31] [32] it uses the frequency of occurrence of words in documents to construct a word-document co-occurrence matrix. Each row of this matrix represents a document or text segment, and each column corresponds to a single word. After pre-processing, the words are converted into vectors in a multi-dimensional space, usually using vectorization techniques such as BoW, TF-IDF or N-grams. These vectors describe the position of each word in a space of several hundred dimensions. Next, LSA applies singular value decomposition (SVD) to the [words x documents] matrix. SVD is a dimensionality reduction method that factorizes the initial matrix X into three components: X= U.Σ.V

- U: a matrix containing the eigenvectors on the left (the "word" vectors).
- Σ: a diagonal matrix containing the largest singular values, representing the importance of latent concepts.
- V: a matrix containing eigenvectors on the right ("document" vectors).

This factorization allows documents and words to be projected into a latent semantic space of reduced dimension. In this space, the similarity between two text units is measured by calculating the cosine of the angle between their respective vectors. This cosine varies between 0 and 1, indicating the semantic proximity between the texts being compared: the closer the value is to 1, the more semantically similar the texts are [26].

Figure 1: Methodology for measuring semantic similarity between documents using LSA

3.2 Neural Network-Based Approaches

3.2.1 Word Embedding-Based Approaches:

These approaches rely on shallow two-layer neural networks to learn dense vector representations of words, capturing their semantic relationships. Semantic similarities between words are represented by the proximity of their vectors in a latent space, meaning that words with similar meanings are located close to each other in this vector space. To generate vector representations of sentences, a common technique is to compute the sum or the average of the vectors representing the different words comprising the sentence. The most widely used word embedding model is Word2Vec, developed by Mikolov and colleagues at Google [33]. Word2Vec can be trained using two main architectures: CBOW (Continuous Bag of Words) and Skip-gram [30][34][35] . The CBOW architecture aims to predict a target word based on the surrounding words within a given context window, while Skip-gram attempts to predict context words from a target word.

3.2.2 Recurrent Neural Network-Based Approaches (RNN):

Recurrent Neural Networks (RNN) have found wide applications in the field of Natural Language Processing (NLP) for modeling data sequences, such as sentences or documents. Unlike traditional neural networks, in an RNN, each neuron receives not only the output from the previous layer but also its own output from a previous time step [36]. This architecture enables RNNs to retain past information, thereby integrating contextual semantics into data processing. However, traditional RNNs struggle to capture information over long sequences, often leading to a loss of information. To overcome this limitation, [37] proposed the Long Short-Term Memory (LSTM) model, which enhances RNNs by better retaining information over longer sequences. This model allows for capturing the semantic properties of text by accommodating sentences of varying lengths and long-range relationships between words within a sentence [38].

3.2.3 Attention-based Approaches:

These approaches leverage attention mechanisms, which can capture long-range relationships between words in a sentence. Attention enables the model to capture dependencies between words and assess their significance for the prediction task, even over long distances. It calculates the significance of each word within a sentence and determines which words to focus on. These architectural advancements rely on the use of multiple self-attention components, each capturing specific relationships between words. Architectures that rely exclusively on attention mechanisms are called Transformers. They were introduced in [39] and are composed of an Encoder-Decoder structure. The encoder consists of a stack of encoders that create a vector representation of a word sequence, while the decoder generates a word sequence from this representation. Each encoder and decoder include, among other components, a self-attention layer that preserves word interdependence within a sequence [40] [41].

The Transformer models have been widely adopted in solving NLP-related tasks due to their ability to generate rich phrase embeddings that capture both the context and semantics of phrases. The most notable Transformer Model is BERT (Bidirectional Encoder Representations from Transformers) developed by Google and primarily used for language understanding tasks [42]. Building upon BERT, more specialized variants have emerged to address nuanced tasks requiring semantic understanding. Notably, Sentence Bert (SBERT) has been developed, employing siamese and triplet network structures to generate semantically meaningful sentence embeddings [43].

4. PROPOSED APPROACH

Building on insights from the literature and considering the scope of our study, which involves conducting a self-assessment of the current state of Moroccan SME based on open-ended responses, we aim for high precision in processing and analyzing these inherently varied and unbounded responses to guide strategic decisions regarding projects and investments. To achieve this, we utilized a multilingual variant of Sentence Transformers, SBERT (paraphrase-multilingual-MiniLM-L12-v2), obtained from the Hugging Face platform [44]. This choice is justified by the model's multilingual capability and its effectiveness in ensuring robust semantic similarity evaluation in a French-language context, which aligns with the Moroccan SMEs needs. Additionally, in the absence of specific training data, our approach leverages reference responses associated with each question. However, since these reference responses are often brief (e.g., yes, no, difficult, very difficult) in audit questionnaires, we enriched them by incorporating keywords extracted from the corresponding questions. This enrichment step provides additional context to the model, thereby enhancing its capacity to perform precise semantic comparisons through the cosine similarity of the generated embedding vectors. The flowchart of the proposed approach is briefly presented in Fig. 2 and consists of four phases:

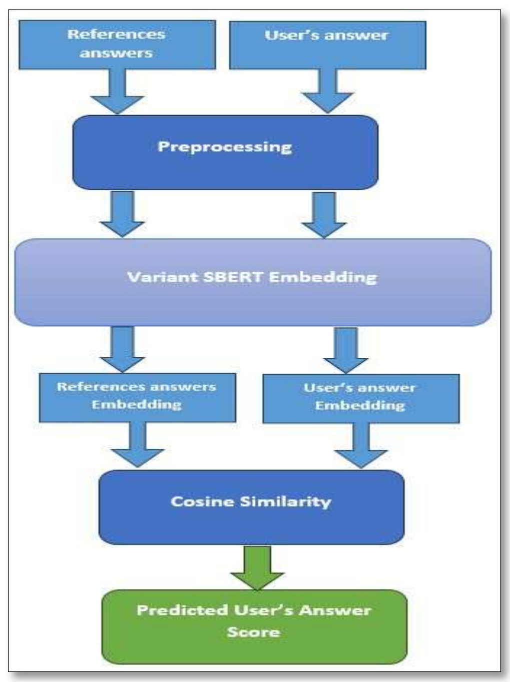

Figure2: Proposed methodology

1. Data Preprocessing: The pre-processing step is designed to prepare both reference responses and open-ended responses for subsequent analysis, ensuring optimal feature extraction for textual similarity algorithms. This step transforms raw, unprocessed text into structured and meaningful data. The sub-steps are as follows:

- Lowercasing: The goal is to normalize the answers by converting all words to lowercase. This ensures that words like" Marketing" and" marketing" are not treated as different entities.
- Removing Special Characters and Numbers: Texts often contain special characters (punctuation, symbols) or numbers that do not add semantic value. These elements are removed to focus solely on relevant words.
- Removing Stopwords: Stopwords are very common words (such as" the"," and"," of" in English) that typically do not add significant semantic value. These words are removed to minimize noise in the data.
- Tokenization: Split the text into separate tokens (words) to prepare the data for embedding.
- Enrichment of Reference Responses: Since reference responses in audit questionnaires are often succinct (e.g., "yes," "no," "difficult"), this step specifically enhances their informativeness. Keywords extracted from the corresponding question are appended to these reference responses, providing additional context and improving the model's ability to calculate semantic similarity accurately. This sub-step applies exclusively to reference responses, as open-ended responses are already detailed by nature.

2. Embedding Generation:

   In this step, we use the SBERT model variant (paraphrase-multilingual-MiniLM-L12-v2) to generate contextual embeddings for the preprocessed answers. Unlike standard Transformer models, which typically process and compare sentences token by token, SBERT fine-tunes the Transformer architecture to produce embeddings that capture the overall semantic meaning of entire sentences [43][45]. While standard Transformers excel at token-level tasks by generating contextual representations for individual tokens, they are less optimized for directly comparing whole sentences. SBERT addresses this limitation by employing Siamese or triplet network structures, enabling it to create single, dense embeddings for sentences. These embeddings can then be efficiently compared using metrics like cosine similarity, making SBERT particularly well-suited for tasks that require sentence-level semantic analysis. This design allows SBERT and similar Sentence Transformer models to surpass the limitations of standard Transformer models in tasks involving semantic comparison of sentences.

3. Similarities calculation:

   - After generating the contextual embeddings for each response in the previous step, the next step involves comparing the embedding of the user's open response with the embeddings of each reference response using cosine similarity. Cosine similarity measures the semantic closeness between two vectors in a continuous vector space, where the embedding represents the semantic meaning of the responses.
   - The cosine similarity between the embedding vectors is calculated using the following formula:

     $$\text{Cosine Similarity} = \frac{A.B}{||A||\,||B||},$$

     where A and B are the embedding vectors of the user's response and the reference response, respectively. The resulting value will range between -1 and 1, where a value closer to 1 indicates a higher degree of semantic similarity between the two responses.

4. Score Assignment:

   - After calculating the cosine similarities, the final task is to assign a score to the user's open answer based on these similarity results. The predicted score is calculated by multiplying the maximum similarity score by the pre-assigned score of the corresponding reference answer. This approach ensures that the user's response is graded relative to the reference that it most closely resembles in terms of semantic content, maintaining the integrity and fairness of the scoring system. The higher the semantic similarity between the user's response and the top reference, the higher the score will be, reflecting the closeness in meaning.

Journal of Theoretical and Applied Information Technology
31st March 2025. Vol.103. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. PROOF OF CONCEPT

To demonstrate the effectiveness of the proposed approach in the previous section and its ability to approximate the results obtained through the manual method—considered ideal due to its precision and the quality of analysis it provides through human expertise in evaluating open-ended responses—we will compare the results of the strategic audit in terms of the ranking of strategic objectives' scores. These scores, derived from our approach based on the multilingual variant of the Sentence Transformer model SBERT (paraphrase-multilingual-MiniLM-L12-v2), will be compared with those of the reference method (results from the manual strategic audit) as well as other automated statistical methods based on the TF-IDF technique and Latent Semantic Analysis (LSA).

To achieve this, a case study was conducted on a Moroccan SME (name undisclosed), which had previously undertaken a manual audit led by a consulting firm. The responses obtained during this manual audit will subsequently be analyzed automatically using our Sentence Transformer-based approach, along with statistical methods such as TF-IDF and LSA.

5.1 Description of the Audit Model (conducted by the auditor)

The strategic audit model utilized by the audit firm for its mission is based on a combination of three complementary analyses:

- SWOT Analysis: To identify the strengths, weaknesses, opportunities, and threats, both internal and external, that influence the company's strategic priorities.
- PESTEL Analysis: To examine the political, economic, sociocultural, technological, environmental, and legal factors affecting the external environment.
- PORTER's Five Forces: To assess competitive dynamics and market pressures.

This model is implemented through a structured questionnaire consisting of 70 open-ended questions, developed based on three analytical frameworks: SWOT, PESTEL, and Porter's model.

Objectif Stratégique	Question	Pourcentag	Réponse R	Score R	Réponse Ref 2	Score Ref 2	Réponse	Score	Réponse Ouverte	Score_Auto
Améliorer l'efficacité opérationnelle	Utilisez-vous des outils numériques pour automatiser l	0,2	Oui	1,00	Partiellement	0,50	Non	0	Nous ne disposons pas encore d'outils numériques pour autor	0,00
Améliorer l'efficacité opérationnelle	Quelle est l'importance des données dans votre prise d	0,2	Très importar	1	Moyenne	0,50	Faible	0	Les données sont essentielles pour nous ; elles guident toutes	1,00
Améliorer l'efficacité opérationnelle	Avez-vous une stratégie numérique pour améliorer la c	0,15	Oui	1	en cours	0,50	non	0	Nous n'avons pas encore développé de stratégie numérique sj	0,50
Accroître la satisfaction client	Disposez-vous d'une stratégie numérique pour amélior	0,25	Oui	1	En cours de développ	0,50	Non	0	Nous n'avons pas encore de stratégie numérique pour amélio	0,00
Accroître la satisfaction client	Utilisez-vous des canaux digitaux pour interagir avec v	0,2	Toujours	1	Parfois	0,50	Rarement	0	Nous utilisons principalement les réseaux sociaux, les newsle	1,00
Accroître la satisfaction client	Avez-vous une approche de personnalisation de l'offre l	0,2	Oui	1	Partiellement	0,50	Non	0	Nous n'avons pas encore d'approche de personnalisation basé	0,00
Renforcer l'innovation et l'agilité	Disposez-vous d'une feuille de route numérique pour le	0,2	Oui	1	en cours	0,50	non	0	Oui, nous avons défini une feuille de route pour les deux proch	1,00
Renforcer l'innovation et l'agilité	Encouragez-vous l'innovation numérique à travers des l	0,15	Régulièremer	1	Occasionnellement	0,50	Jamais	0	Nous n'avons pas mis en place de telles initiatives pour encou	0,00
Renforcer l'innovation et l'agilité	Utilisez-vous des technologies numériques pour accél	0,15	Oui	1	Partiellement	0,50	Non	0	Nous utilisons des outils de gestion de projets agiles et on se l	0,50
Renforcer l'innovation et l'agilité	Avez-vous adopté des méthodes agiles pour le dévelop	0,15	Oui	1	En partie	0,50	Non	0	Nous avons adopté les méthodes agiles pour développer rapic	1,00
Accroître la satisfaction client	Avez-vous une stratégie de contenu numérique pour en	0,15	Oui	1	Parfois	0,50	Rarement	0	Nous avons une stratégie de contenu pour engager activemen	1,00
Améliorer l'efficacité opérationnelle	Disposez - vous d'outils d'analyse pour évaluer l'efficac	0,15	Oui	1	Partiellement	0,50	Non	0	Nous utilisons des outils d'analyse pour mesurer l'efficacité de	0,50
Accroître la satisfaction client	Utilisez-vous des outils d'IA pour améliorer le service c	0,2	Oui	1	Partiellement	0,50	Non	0	Nous utilisons des chatbots IA pour répondre plus rapidement	1,00
Renforcer l'innovation et l'agilité	Investissez-vous régulièrement dans la recherche et dé	0,15	Oui	1	Occasionnellement	0,50	Non	0	Oui, nous investissons dans des projets de recherche pour fav	1,00
Renforcer l'innovation et l'agilité	Avez-vous mis en place des partenariats externes pour	0,1	Oui	1	Parfois	0,50	Non	0	Nous collaborons des fois avec des start-ups pour stimuler nu	0,50
Améliorer l'efficacité opérationnelle	Suivez-vous des indicateurs de performance numériqu	0,1	Oui	1	Partiellement	0,50	Non	0	Nous surveillons en permanence les indicateurs de performar	1,00
Améliorer l'efficacité opérationnelle	Avez-vous un programme de formation numérique pou	0,1	Oui	1	En cours	0,50	Non	0	Nous avons un programme de formation complet pour amélio	1,00
Renforcer l'innovation et l'agilité	Avez-vous des politiques de flexibilité pour tester rapid	0,1	Oui	1	En cours	0,50	Non	0	Nous avons des processus flexibles pour tester de nouvelles ti	0,50
Améliorer l'efficacité opérationnelle	Disposez-vous de systèmes automatisés pour la gestio	0,1	Oui	1	Partiellement	0,50	Non	0	Nous utilisons des systèmes automatisés pour gérer efficacen	0,50
Optimiser la gestion des ressources	Avez-vous mis en place une gestion numérique des res	0,3	Oui	1	Partiellement	0,50	Non	0	Nous avons mis en place une solution numérique de gestion d	0,50
Optimiser la gestion des ressources	Utilisez-vous des outils d'analyse pour optimiser l'alloc	0,25	Oui	1	Partiellement	0,50	Non	0	Oui, nous utilisons des outils d'analyse de données pour optin	1,00
Optimiser la gestion des ressources	Disposez-vous de tableaux de bord pour suivre l'utilisat	0,2	Oui	1	Partiellement	0,50	Non	0	Nous manquons actuellement d'outils de tableau de bord en t	0,00
Optimiser la gestion des ressources	Automatisez-vous la gestion des stocks et des fournitu	0,15	Oui	1	Partiellement	0,50	Non	0	Nous utilisons un système d'automatisation pour surveiller le	1,00
Optimiser la gestion des ressources	Mettez-vous à jour régulièrement vos systèmes de gest	0,1	Oui	1	Partiellement	0,50	Non	0	Nos systèmes de gestion sont mis à jour de façon sporadique.	0,50
Augmenter la sécurité des systèmes d'I	Avez-vous une politique de cybersécurité bien définie ?	0,25	Oui	1	Partiellement	0,50	Non	0	Oui, nous avons une politique de cybersécurité en place, revu	1,00
Augmenter la sécurité des systèmes d'I	Disposez-vous d'une équipe dédiée à la gestion des inc	0,25	Oui	1	Partiellement	0,50	Non	0	Nous avons une équipe interne dédiée à la gestion des incider	1,00
Augmenter la sécurité des systèmes d'I	Utilisez-vous des technologies de chiffrement pour pro	0,2	Oui	1	Partiellement	0,50	Non	0	Toutes nos données sensibles sont protégées par des technol	1,00
Augmenter la sécurité des systèmes d'I	Effectuez-vous des audits réguliers de sécurité informa	0,15	Oui	1	Partiellement	0,50	Non	0	Nos audits de sécurité informatique sont effectués d'une faço	0,50
Augmenter la sécurité des systèmes d'I	Avez-vous implémenté une authentification multifacto	0,15	Oui	1	Partiellement	0,50	Non	0	Nous avons commencé à implémenter la MFA pour les accès	0,50

Figure3: Strategic audit questions

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

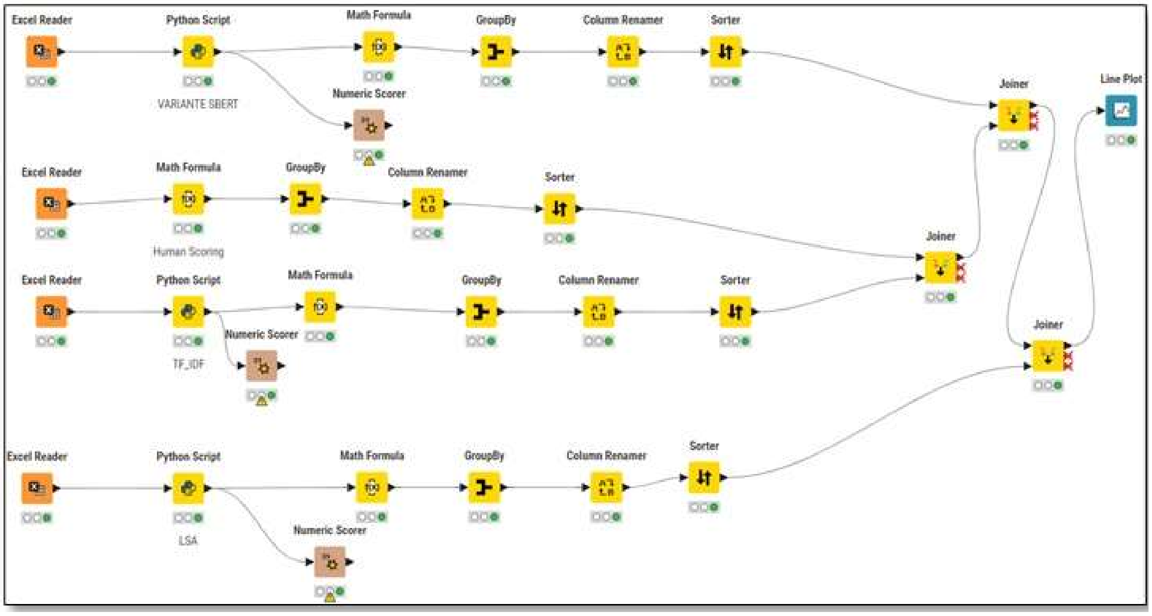

Figure4: Workflows Using SBERT, TF-IDF and LSA approaches

The questionnaire is structured as follows (see fig. 3):

- Each question, indexed as k, has a direct impact on a strategic objective i.
- Each objective is linked to one or more questions, with a defined impact p_{ik} for each question k.
- This impact is measured as a percentage p_{ik}, indicating the relative importance of question k to objective i.
- Each question is also associated with a set of annotated reference responses, serving as a baseline for evaluating open-ended responses.
- The score for each strategic objective i is calculated by the auditor by weighing the responses to the questions based on their importance p_{ik} using the following formula: Score (strategic objective i) = $1/N_i \sum_k p_{ik} \times SCORE\_Q_k$

  Where Ni is the total number of questions impacting the strategic objective i.

5.2 Evaluation

The evaluation part aims to compare our proposed approach, which leverages a variant of SBERT (paraphrase-multilingual-MiniLM-L12-v2) along with statistical techniques, to the manual method in terms of performance and accuracy. The open-ended responses obtained through the manual method were analyzed using a series of algorithms to assess their effectiveness.

Three different algorithms were selected for this analysis. The first algorithm employs the statistical technique of TF-IDF, the second utilizes Latent Semantic Analysis (LSA), and the third is our proposed approach, which is based on a variant of SBERT (paraphrase-multilingual-MiniLM-L12-v2). These algorithms were implemented using KNIME [46], a data analysis platform that facilitates the creation of workflows for automating analyses. To ensure clarity and reproducibility, each method was developed and executed within a separate workflow. The process begins with importing the data from an Excel file using the Excel Reader node. Next, Python Script nodes were used in each workflow to execute the algorithms associated with the three approaches (SBERT variant, LSA, and TF-IDF). Finally, mathematical calculations and aggregation of results were performed (using Math Formula, GroupBy nodes, etc.) to obtain the final scores for the strategic objectives (see fig. 4).

6. RESULTS

To evaluate the performance of each approach, we connected the output of the Python Script node in each workflow to a node called Numeric Scorer. This node generates a metrics table that includes measures such as MSE (Mean Squared Error), MAE (Mean Absolute Error), the coefficient of determination (R^2), and other statistical measures to assess the quality of predictions. The Numeric Scorer is configured to compare two columns: the prediction column, which contains the scores calculated by the algorithm for each approach (SBERT variant, LSA, or TF-IDF), and the reference column, which corresponds to the score assigned by the auditor.

The results obtained in table 2 demonstrate a clear difference in performance between the three approaches for calculating the scores of open responses. Our method based on the SBERT variant stands out as the most effective method, achieving a

Mean Absolute Error (MAE) of 0.24, a Mean Squared Error (MSE) of 0.1, and a Root Mean Squared Error (RMSE) of 0.31, reflecting a higher accuracy compared to the other algorithms.

Additionally, the Mean Signed Difference (MSD) of -0.23 indicates a slight tendency of our method to overestimate actual values, although this overestimation remains relatively minor.

In comparison, the LSA and TF-IDF algorithms yielded significantly larger errors, with higher MAE, MSE, and RMSE values, indicating less accurate predictions. Their negative MSDs of -0.53 for LSA and -0.62 for TF-IDF reveal an even stronger tendency to overestimate actual values.

Furthermore, Joiner nodes were used to create a summary table of the strategic objective scores obtained from the three approaches, along with the manual scores assigned by the auditor, as illustrated in fig. 5. This table was then connected to the Line Plot node, allowing the display of four curves: the strategic objective scores according to the three approaches and the manual scores from the auditor (see fig. 6).

According to the graph in fig. 6, we observe that the approach based on the SBERT variant (blue line) and the one based on human evaluation (green line) follow similar trends, suggesting that both approaches classify the strategic objectives almost identically. In other words, the proposed approach using SBERT successfully reproduces human judgment in ranking the strategic objectives, demonstrating that this natural language processing algorithm effectively captures the priorities or relative importance of each objective. Both curves display a parallel decline, indicating that for each strategic objective, our proposed approach assigns scores comparable to those of human evaluations. In contrast, the LSA approach shows significant fluctuations, with a strong increase for the second objective followed by a marked decline, indicating a divergence from human classification. On the other hand, TF-IDF assigns very low and consistent scores, showing little variation between the objectives, suggesting that it fails to effectively differentiate between them.

While our findings demonstrate the effectiveness of the SBERT-based approach in aligning with human evaluations, certain limitations should be acknowledged. The observed slight overestimation bias (MSD of -0.23) suggests that the model may benefit from further calibration to enhance accuracy. Additionally, the reliance on predefined reference responses could introduce biases, particularly if these responses do not encompass the full diversity of potential answers. Another key consideration is the variability in results across different strategic contexts—while the current evaluation confirms strong performance within our dataset, broader applicability across industries remains to be explored. Addressing these aspects through dataset expansion and adaptive scoring mechanisms would further strengthen the robustness of our method.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Performance Metrics for Each Approach

	SBERT Variant	LSA	TF - IDF
Mean absolute error	0.24	0.53	0.63
Mean squared error	0.1	0.45	0.52
Root mean squared error	0.31	0.67	0.72
Mean signed difference	-0.23	-0.53	-0.62


Rows: 5 | Columns: 5
Table Statistics

#	RowID	Objectif Stratégique (String)	Score SBERT Variant (Number (double))	Auditor Scoring (Number (double))	Score TF IDF (Number (double))	Score LSA (Number (double))
1	Row...	Augmenter la sécurité des systèmes d'info...	0.104	0.17	0.008	0
2	Row...	Optimiser la gestion des ressources	0.077	0.12	0.006	0.05
3	Row...	Accroître la satisfaction client	0.076	0.11	0.002	0
4	Row...	Renforcer l'innovation et l'agilité	0.074	0.096	0.008	0.021
5	Row...	Améliorer l'efficacité opérationnelle	0.053	0.086	0.002	0.039

Figure5: Strategic objective scores for each approach

Figure 6: Visual Comparison of Strategic Objective Scoring for each approach

7. CONCLUSION

This study proposed a novel AI-driven self-assessment method for evaluating SMEs' current state using open-ended responses in the context of digital transformation. By leveraging a multilingual Sentence Transformers model, our approach provides an accessible and cost-effective alternative to traditional expert assessments. The experimental results confirmed its effectiveness, as it closely aligns with human evaluations, achieving the lowest MAE, MSE, and RMSE scores compared to LSA and TF-IDF.

However, despite these promising results, some limitations must be acknowledged. First, our method relies on reference responses, which may introduce bias if they do not adequately cover the diversity of possible answers. Second, while SBERT demonstrates strong alignment with human assessments, our findings indicate a slight tendency to overestimate scores, highlighting the need for additional calibration mechanisms. Third, the current approach does not incorporate contextual variations across different industries and strategic environments, which may affect its generalizability.

To address these challenges, future research should focus on expanding the dataset by collecting more open-ended responses from SMEs in diverse business contexts, allowing for better fine-tuning of the model. Additionally, exploring hybrid models that combine deep learning with rule-based adjustments could improve both accuracy and interpretability. Extending this approach to other types of audits, such as digital or cultural maturity assessments, could contribute to the development of a more comprehensive self-assessment framework for organizations.

Beyond textual analysis, an open research challenge is the integration of multimodal assessment approaches, such as speech-based responses. Allowing SMEs to provide verbal feedback could enrich input data while making the assessment more accessible to users with limited writing skills. This enhancement could open new perspectives on how AI refines decision-making processes and advances self-assessment methodologies.

In conclusion, while our approach demonstrates strong potential in automating strategic audit evaluations, continuous refinement and expanded datasets are crucial to enhancing its applicability across various industries. Future work should further explore its adaptability to different organizational assessments, paving the way for a more holistic and scalable self-evaluation framework.

REFERENCES :

[1] F. Khanboubi and A. Boulmakoul, "Etat de l'art sur la transformation digitale : focus sur le domaine bancaire".
[2] J. Paasi, "Towards a new era in manufacturing: Final report of VTT's For Industry spearhead programme," 2017.
[3] P. C. Verhoef *et al.*, "Digital transformation: A multidisciplinary reflection and research agenda," *Journal of Business Research*, vol. 122, pp. 889–901, Jan. 2021, doi: 10.1016/j.jbusres.2019.09.022.
[4] P. Besson and F. Rowe, "Strategizing information systems-enabled organizational transformation: A transdisciplinary review and new directions," *The Journal of Strategic Information Systems*, vol. 21, no. 2, Art. no. 2, Jun. 2012, doi: 10.1016/j.jsis.2012.05.001.
[5] K. Crowston and M. D. Myers, "Information technology and the transformation of industries: three research perspectives," *The Journal of Strategic Information Systems*, vol. 13, no. 1, Art. no. 1, Mar. 2004, doi: 10.1016/j.jsis.2004.02.001.
[6] N. Venkatraman, "IT-enabled business transformation: from automation to business scope redefinition," *MIT Sloan Management Review*, vol. 35, no. 2, Art. no. 2, 1994.
[7] Q. T. Pham, "Measuring The ICT Maturity Of SMEs".
[8] J. Kääriäinen, L. Saari, M. Tihinen, S. Perätalo, and T. Koivumäki, "Supporting the digital transformation of SMEs — trained digital evangelists facilitating the positioning phase," *IJISPM*, vol. 11, no. 1, Art. no. 1, Apr. 2023, doi: 10.12821/ijispm110101.
[9] O. Kuusisto, J. Kääriäinen, K. Hänninen, and M. Saarela, "Towards a Micro-Enterprise–Focused Digital Maturity Framework:," *International Journal of Innovation in the Digital Economy*, vol. 12, no. 1, Art. no. 1, Jan. 2021, doi: 10.4018/IJIDE.2021010105.
[10] J. Kääriäinen, P. Pussinen, L. Saari, O. Kuusisto, M. Saarela, and K. Hänninen, "Applying the positioning phase of the digital transformation model in practice for SMEs: toward systematic development of digitalization," *IJISPM*, vol. 8, no. 4, Art. no. 4, Sep. 2021, doi: 10.12821/ijispm080402.
[11] P. Parviainen, M. Tihinen, J. Kääriäinen, and S. Teppola, "Tackling the digitalization challenge: how to benefit from digitalization in practice," *International Journal of Information Systems and Project Management*, vol. 5, no. 1, Art. no. 1, 2017, doi: 10.12821/ijispm050104.
[12] V. Stich, V. Zeller, J. Hicking, and A. Kraut, "Measures for a successful digital transformation of SMEs," *Procedia CIRP*, vol. 93, pp. 286–291, 2020, doi: 10.1016/j.procir.2020.03.023.
[13] F. Pirola, C. Cimini, and R. Pinto, "Digital readiness assessment of Italian SMEs: a case-study research," *JMTM*, vol. 31, no. 5, Art. no. 5, Nov. 2019, doi: 10.1108/JMTM-09-2018-0305.
[14] J. Soluk and N. Kammerlander, "Digital transformation in family-owned Mittelstand firms: A dynamic capabilities perspective," *European Journal of Information Systems*, vol. 30, no. 6, Art. no. 6, Nov. 2021, doi: 10.1080/0960085X.2020.1857666.
[15] D. Ulas, "Digital Transformation Process and SMEs," *Procedia Computer Science*, vol. 158, pp. 662–671, 2019, doi: 10.1016/j.procs.2019.09.101.
[16] R. C. Santos and J. L. Martinho, "An Industry 4.0 maturity model proposal," *JMTM*, vol. 31, no. 5, Art. no. 5, Dec. 2019, doi: 10.1108/JMTM-09-2018-0284.
[17] T. Haryanti, N. A. Rakhmawati, and A. P. Subriadi, "The Extended Digital Maturity Model," *BDCC*, vol. 7, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/bdcc7010017.

[18] L. Saari, O. Kuusisto, and J. Häikiö, *ManuMaturity: A maturity tool for manufacturing companies to reach beyond Industry 4.0*. VTT Technical Research Centre of Finland, 2021.
[19] A. Amalia, D. Gunawan, Y. Fithri, and I. Aulia, "Automated Bahasa Indonesia essay evaluation with latent semantic analysis," *J. Phys.: Conf. Ser.*, vol. 1235, no. 1, Art. no. 1, Jun. 2019, doi: 10.1088/1742-6596/1235/1/012100.
[20] L. Zhang, Y. Huang, X. Yang, S. Yu, and F. Zhuang, "An automatic short-answer grading model for semi-open-ended questions," *Interactive Learning Environments*, vol. 30, no. 1, Art. no. 1, Jan. 2022, doi: 10.1080/10494820.2019.1648300.
[21] A. Condor and M. Litster, "Automatic short answer grading with SBERT on out-of-sample questions," 2021.
[22] D. Wilianto and A. Suganda Girsang, "Automatic Short Answer Grading onHigh School's E-Learning Using Semantic Similarity Methods," *TEM Journal*, pp. 297–302, Feb. 2023, doi: 10.18421/TEM121-37.
[23] J. Xue, X. Tang, and L. Zheng, "A Hierarchical BERT-Based Transfer Learning Approach for Multi-Dimensional Essay Scoring," *IEEE Access*, vol. 9, pp. 125403–125415, 2021, doi: 10.1109/ACCESS.2021.3110683.
[24] A. Kanavos, I. Karamitsos, A. Mohasseb, and V. C. Gerogiannis, "Comparative Study of Machine Learning Algorithms and Text Vectorization Methods for Fake News Detection," in *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Volos, Greece: IEEE, Jul. 2023, pp. 1–8. doi: 10.1109/IISA59645.2023.10345953.
[25] T. Hasan and A. Matin, "Extract Sentiment from Customer Reviews: A Better Approach of TF-IDF and BOW-Based Text Classification Using N-Gram Technique," in *Proceedings of International Joint Conference on Advances in Computational Intelligence*, M. S. Uddin and J. C. Bansal, Eds., in Algorithms for Intelligent Systems. , Singapore: Springer Singapore, 2021, pp. 231–244. doi: 10.1007/978-981-16-0586-4_19.
[26] S. Nayak, R. Agarwal, S. K. Khatri, and M. Mohammadian, "Student Outcome Assessment on Structured Query Language using Rubrics and Automated Feedback Generation," *IJACSA*, vol. 15, no. 3, Art. no. 3, 2024, doi: 10.14569/IJACSA.2024.0150374.
[27] S. T. Dumais, G. W. Furnas, T. K. Landauer, S. Deerwester, and R. Harshman, "Using latent semantic analysis to improve access to textual information," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '88*, Washington, D.C., United States: ACM Press, 1988, pp. 281–285. doi: 10.1145/57167.57214.
[28] T. K. Landauer and S. T. Dutnais, "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge".
[29] Y. Kalmukov, "Comparison of Latent Semantic Analysis and Vector Space Model for Automatic Identification of Competent Reviewers to Evaluate Papers," *IJACSA*, vol. 13, no. 2, Art. no. 2, 2022, doi: 10.14569/IJACSA.2022.0130209.
[30] O. Ahmia, N. Béchet, P.-F. Marteau, and A. Garel, "Utilité d'un couplage entre Word2Vec et une analyse sémantique latente : expérimentation en catégorisation de données textuelles.".
[31] R. M. Suleman and I. Korkontzelos, "Extending latent semantic analysis to manage its syntactic blindness," *Expert Systems with Applications*, vol. 165, p. 114130, Mar. 2021, doi: 10.1016/j.eswa.2020.114130.
[32] A. A. Nafea, N. Omar, and Z. M. Al-qfail, "Artificial Neural Network and Latent Semantic Analysis for Adverse Drug Reaction Detection," *Baghdad Sci.J*, May 2023, doi: 10.21123/bsj.2023.7988.
[33] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Sep. 06, 2013, *arXiv*: arXiv:1301.3781. Accessed: Oct. 04, 2024. [Online]. Available: http://arxiv.org/abs/1301.3781
[34] D. Ramesh and S. K. Sanampudi, "An automated essay scoring systems: a systematic literature review," *Artif Intell Rev*, vol. 55, no. 3, Art. no. 3, Mar. 2022, doi: 10.1007/s10462-021-10068-2.
[35] K. W. Trisna, J. Huang, and H. Lei, "From context-independent embedding to transformer: exploring sentiment classification in online reviews with deep learning approaches," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 19, 2024.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[36] Z. He, C. E. Dumdumaya, and V. A. Quimno, "MEASUREMENT OF SEMANTIC TEXT SIMILARITY," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 5, 2024.

[37] H. Palangi *et al.*, "Semantic Modelling with Long-Short-Term Memory for Information Retrieval," Feb. 27, 2015, *arXiv*: arXiv:1412.6629. Accessed: Oct. 04, 2024. [Online]. Available: http://arxiv.org/abs/1412.6629

[38] A. K. Baareh *et al.*, "predicting daily river flow using long short- term memory (lstm): a deep learning approach," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 23, 2024.

[39] S. Haller, A. Aldea, C. Seifert, and N. Strisciuglio, "Survey on Automated Short Answer Grading with Deep Learning: from Word Embeddings to Transformers," Mar. 11, 2022, *arXiv*: arXiv:2204.03503. Accessed: May 05, 2024. [Online]. Available: http://arxiv.org/abs/2204.03503

[40] S. Ludwig, C. Mayer, C. Hansen, K. Eilers, and S. Brandt, "Automated Essay Scoring Using Transformer Models," *Psych*, vol. 3, no. 4, Art. no. 4, Dec. 2021, doi: 10.3390/psych3040056.

[41] "Automatic Short Answer Grading Using Deep Learning.pdf."

[42] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 24, 2019, *arXiv*: arXiv:1810.04805. Accessed: Oct. 04, 2024. [Online]. Available: http://arxiv.org/abs/1810.04805

[43] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, 2019, pp. 3980–3990. doi: 10.18653/v1/D19-1410.

[44] Hugging Face, "Paraphrase-multilingual-MiniLM-L12-v2 Model Card." Accessed: Jun. 05, 2024. [Online]. Available: https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2

[45] A. Condor and M. Litster, "Automatic short answer grading with SBERT on out-of-sample questions," 2021.

[46] "KNIME Analytics Platform," KNIME | Open for Innovation. Accessed: Aug. 02, 2024. [Online]. Available: https://www.knime.com/knime-analytics-platform