NEBULA Framework: An Adaptive Framework for Unstructured Description to Solve Cold Start Problem

I Gusti Agung Gede Arya Kadyanan[1*], Ni Made Ary Esta Dewi Wirastuti[2],
Gede Sukadarmika[3], Ngurah Agus Sanjaya ER[4]
Doctoral Program Engineering Science-Faculty of Technic, Udayana University, Badung, Indonesia[1]
Department of Electro-Faculty of Technic, Udayana University, Badung, Indonesia[2, 3]
Department of Informatics-Faculty of Mathematics and Natural Science, Udayana University, Badung, Indonesia[4]

***Abstract*—The cold start problem is one of the main challenges in recommendation systems, especially when the system has to provide recommendations for new items that do not yet have a history of interaction. Although various approaches have been developed, most still use conventional interaction-based methods, which are not optimal in providing accurate recommendations for new items that only have minimal and unstructured descriptive information. This research aims to provide recommendations for new items that lack interaction history and have unstructured descriptive information by addressing the cold start problem more adaptively. The proposed model is based on Named Entity Recognition (NER) and metadata representation as an adaptive framework capable of adjusting recommendation methods based on the availability of initial information. For new items, the system utilizes basic attributes such as product type, materials, and origin, and employs an adaptive approach for rating prediction. Testing results demonstrate system performance with an Accuracy of 0.967, Precision of 0.838, Recall of 0.846, F1-score of 0.842, and an average Mean Absolute Error (MAE) of 0.159. This adaptive framework proved to be superior to conventional approaches, with improvements in Precision of 15.59%, Recall of 17.50%, F1-score of 16.54%, and a significant reduction in MAE. Additionally, the Kappa value of 0.69 indicates a high level of agreement (substantial agreement) among validators. These findings demonstrate that the system is not only more accurate in recommending new items but also more reliable under minimal data conditions, thereby enhancing user confidence. Overall, this NER and metadata-based framework can serve as an effective solution for addressing the cold start problem and improving recommendation quality during the initial stages.**

***Keywords*—Cold start; adaptive framework; recommender system; NER; unstructured description**

I. INTRODUCTION

Recommendation systems have become an important component in various modern digital platforms, including e-commerce, media services, tourism, and culture. These systems help users filter information and find relevant products or services based on personal preferences, thereby improving the efficiency of interactions and overall user satisfaction [1], [2], [3].

However, one of the main challenges in recommendation systems is the cold start problem, which occurs when new items or users lack sufficient historical data, such as ratings or interaction history. This situation makes it difficult for traditional recommendation algorithms, which generally rely on previous interaction patterns [4]. The cold start problem can arise in various scenarios, such as when new products are added to the system, new users join, or items have limited descriptive data [5], [6], [7].

The cold start problem is particularly critical because it directly impacts the exposure and accessibility of new products. In commercial sectors such as consumer goods, fashion, technology, or daily necessities, new items often fail to receive optimal recommendations despite having high market potential. Existing recommendation systems still heavily rely on historical interaction data, causing new products to be less visible to users [8]. This situation highlights a research gap: the need for an approach that can leverage descriptive information or metadata more adaptively, rather than solely relying on interaction history.

Previous studies have attempted to address the cold start problem using different approaches. For example, a hybrid collaborative filtering method combined with content-based filtering [9], or the use of NLP-based descriptive representations [10], [11]. However, most of these studies still face limitations in terms of metadata quality and difficulties in extracting information from minimal and unstructured product descriptions. Additionally, existing models are generally rigid and not fully adaptive to data variations in specific domains, resulting in system performance degradation when faced with new products with incomplete or inconsistent metadata. This highlights a research gap, namely the need for a new approach that not only processes metadata but also possesses adaptive, flexible, and contextual capabilities in handling variations in product descriptions.

As a solution, this study proposes a metadata-based recommendation system model utilizing Natural Language Processing (NLP) with a focus on Named Entity Recognition (NER) techniques. This approach enables the system to extract important attributes from product descriptions, such as category, materials, and region of origin, to form vector representations that can be used in similarity calculations between items [12], [13]. To address conditions where no items exceed the similarity threshold, the system is also equipped with a clustering-based fallback strategy, ensuring that recommendations remain relevant. The model's strengths lie not only in its ability to address content-based cold start issues without relying on user interaction history, but also in its

*Corresponding author.

adaptive nature, which can adjust to limitations in descriptions, differences in writing style, and language variations in the domain of traditional crafts.

Thus, the objective of this research is to develop an adaptive framework based on metadata and NER to address the cold start problem for new items, especially when descriptive information is still limited. From a practical standpoint, the system is expected to increase the exposure of new products, such as traditional crafts, thereby contributing to local economic empowerment. From an academic standpoint, this research enriches the literature on NLP and metadata-based recommendation systems and offers a new contribution in the form of a more flexible and contextual adaptive approach, which opens up opportunities for wider application in various domains with similar challenges.

The contributions of this research can be summarized into three main aspects. First, from an algorithmic perspective, the proposed framework integrates multilingual Named Entity Recognition (NER) with adaptive embedding selection (spaCy for English and BERT for Indonesian), dynamic database switching, and clustering-based fallback mechanisms to ensure recommendation continuity under sparse conditions. Second, from a methodological perspective, this study advances the handling of unstructured product metadata by transforming free-form descriptions of traditional crafts into structured semantic entities that can be embedded and compared systematically. Third, from an application perspective, this research demonstrates the applicability of the framework to the domain of traditional crafts, showing how cultural product metadata can be leveraged to overcome cold start limitations while preserving local knowledge in digital marketplaces. Collectively, these contributions distinguish this research from existing NLP and content-based solutions, positioning the framework not merely as an applied refinement but also as a substantial advancement in adaptive cold start handling.

The remainder of this paper is organized as follows. Section II reviews related works that address the cold start problem in recommendation systems. Section III introduces the proposed metadata-based adaptive model with Named Entity Recognition. Section IV presents the model analysis and its components. Section V provides the performance assessment through experimental evaluation. Finally, Section VI concludes the paper and outlines potential directions for future research.

II. Related Work

The cold start problem has become one of the main challenges in the development of modern recommendation systems. Xiao et al. [14] through the UPRec approach, argue that a pre-training process that explicitly integrates user attributes can enrich user representations and reduce dependence on limited historical data, thereby effectively addressing cold start in sequential recommendation scenarios.

A similar approach was also comprehensively reviewed by Zhou et al. [15], who grouped deep learning-based methods such as BERT4Rec and Graph Neural Network (GNN) as cutting-edge solutions to sparsity and cold start problems, through the formation of deeper and more contextual user and item embeddings.

In the content and emotion-based domain, Jiang et al. [16] developed a recommendation system by combining Word2Vec and LSTM for sentiment analysis, demonstrating that even short reviews can be utilized to form user emotion modeling in social recommendation systems. An NLP-based approach was also adopted by Mishra et al. [11], who combined review classification using Random Forest and a collaborative filtering-based filtering model. The study demonstrated that review quality and text analysis can improve recommendation accuracy even under cold start conditions.

On the other hand, content-based filtering approaches relying on product metadata have been extensively developed for domains facing user interaction limitations. Negara et al. [17] built a recommendation system for an NFT marketplace using only item name and description features based on TF-IDF and cosine similarity, which proved capable of delivering relevant results even without any user ratings at all.

A metadata-based approach was also proposed by Lestari et al. [18] in the context of MSME marketplaces. They combined clustering and imputation techniques to fill data gaps in new products and showed that this combination was able to significantly improve prediction accuracy by up to 100% in KNN and Naive Bayes algorithms.

Widayanti et al. [9] also proposed a hybrid filtering-based recommendation system that combines Collaborative Filtering and Content-Based Filtering to overcome cold start and expand the diversity of recommendations. This hybrid strategy leverages the advantages of CF in detecting user preference patterns and the strength of CBF in explicitly extracting content features, thereby providing richer and more contextual predictions.

Panteli et al. [19] offer an approach based on frequent pattern mining and user clustering that can estimate new users preferences through distinctive purchasing patterns within each cluster of existing users. This model has proven successful in addressing the cold start problem, achieving a precision accuracy of 90.3% and doubling data density in experiments with MovieLens.

Gupta et al. [20] developed an adaptive e-commerce recommendation system using a hybrid deep learning-based approach, integrating Multilayer Perceptron (MLP), Matrix Factorization, autoencoder, and transformer-based re-ranking. This approach is highly effective in addressing cold start and data sparsity by leveraging various user behavior sources such as clicks, searches, and explicit ratings, achieving an accuracy performance of up to 91%.

Ramakrishna et al. [21] introduced the Hybrid Collaborative Filtering (HCoF) approach, which combines three filtering approaches: Collaborative, Social, and Semantic. This system successfully addresses the cold start problem on social networks by relying on similarity based on interests, location, and social trust levels. Evaluations show that this approach achieves an F1-score of 0.651 and high fairness across various user segments.

Afoudi et al. [22] combined Collaborative Filtering, Content-Based Filtering, and Self-Organizing Map (SOM) into a single hybrid system architecture. This model strengthens user preference predictions based on genre and demographics and is capable of generating relevant recommendations even under data constraints or cold start conditions.

In addition to technical approaches, Vomberg et al. [23] discuss the cold start problem in the context of AI system implementation strategies that are still in their early stages. In their article, they emphasize the importance of building a data network effect to accelerate the data accumulation process in new systems. This approach focuses on organizational and business strategy aspects, indicating that solutions to the cold start problem are not always algorithmic but can also be achieved through system design that promotes gradual and targeted data growth.

From the various studies reviewed, it appears that the approaches to overcoming cold start problems in recommendation systems are very diverse, ranging from attribute-based pre-training, the use of deep learning embedding, NLP-based sentiment analysis, to hybrid methods that combine Collaborative Filtering (CF) and Content-Based Filtering (CBF). However, each study still has certain limitations, whether related to dependence on historical data, metadata quality, or adaptability to data variations in specific domains. To clarify the position of this research compared to previous works, Table I presents a summary of the comparison of approaches, advantages, disadvantages, and research gaps that this research attempts to fill.

Table I shows that previous studies have made important contributions to overcoming the cold start problem, but most are still limited to dependence on historical data, review quality, or limitations in adapting to product description variations. This study is different because it offers a more adaptive approach through the use of product metadata and Named Entity Recognition (NER) techniques, which enable the system to extract important attributes contextually. Thus, the main contribution of this study is to provide a solution for cold start items in traditional handicraft products, while enriching the literature on metadata-based recommendation systems and NLP in the local cultural domain.

TABLE I. COMPARISON OF PREVIOUS RESEARCH WITH THIS RESEARCH

Researcher & Year	Approach/Method	Strengths	Research Gap	Contribution of This Research
Xiao et al. (2023) [14]	UPRec with pre-training of user attributes	Enriches user representation, reduces reliance on historical data	Focuses on user-based cold start, does not address item cold start	Focus on item-based cold start using product metadata
Zhou et al. (2023) [15]	Deep learning (BERT4Rec, GNN)	More contextual user/item embedding, effective for sparsity	Requires large-scale data, less adaptive in narrow domains	Adaptive approach based on descriptive metadata
Jiang et al. (2020) [16]	Word2Vec + LSTM for sentiment analysis	Utilizes short reviews, captures user emotions	Relies on review quality, not always available	Independent from reviews, focused on product metadata
Mishra et al. (2024) [11]	NLP + Random Forest + Collaborative Filtering	Combines text and interaction, improves accuracy	Still requires interaction data, vulnerable to item cold start	Handles cold start without user interaction data
Negara et al. (2023) [17]	TF-IDF + Cosine Similarity (NFT marketplace)	Relevant even without ratings	Simple metadata, noisy, lacks deeper representation	Richer feature extraction with NER
Lestari et al. (2024) [18]	Metadata + clustering + imputation (MSME marketplace)	Fills missing data, significant accuracy improvement	Non-contextual, limited to specific domains	Adds local and cultural aspects
Widayanti et al. (2023) [9]	Hybrid CF + CBF	Provides higher recommendation diversity	Still relies on user interaction data	Independent from interaction history, adaptive metadata
Panteli et al. (2023) [19]	Frequent pattern mining + user clustering	Estimates new user preferences, precision 90.3%	Focuses on user cold start, not new items	Addresses cold start specifically for new items using descriptions
Gupta et al. (2024) [20]	Deep learning hybrid (MLP, MF, Autoencoder, Transformer)	Highly accurate (91%), leverages multi-source data	Complex, requires large-scale data (clicks, searches, ratings)	Lightweight, focused on descriptive metadata
Ramakrishna et al. (2023) [21]	Hybrid Collaborative Filtering (HCoF): CF + Social + Semantic	Effective in social networks, F1-score 0.651	Context-limited (social networks), still requires interest/location data	Local culture and product context without social interactions
Afoudi et al. (2021) [22]	CF + CBF + Self-Organizing Map (SOM)	Strengthens preference prediction, relevant in limited data scenarios	Complex, still dependent on genre/demographics	Focus on metadata of traditional handicraft products
Vomberg et al. (2023) [23]	AI implementation strategy (data network effect)	Accelerates accumulation of new data through system design	Does not offer algorithmic technical solution	Technical adaptive solution based on NLP (NER + metadata)
This Research	Metadata + NER + embedding + fallback clustering	Adaptive, cultural product focus, no historical data required	–	Tackles item cold start for handicraft products in context

III. PROPOSED MODEL

The proposed recommendation system model consists of three main stages, such as descriptive extraction, similarity estimation, and rating prediction. These three stages are summarized in the block diagram in Fig. 1, where each block shows the processing flow from new item input to rating estimation. This process is designed adaptively to address the cold start problem for new items that have no interaction history or user ratings [24]. Suppose a new item is denoted as

N, and the set of existing items in the database is denoted as $C = \{C_1, C_2, \dots, C_j\}$, where each C_j represents a product that already has descriptive information and ratings from users. The framework used in this study is named NEBULA (Named Entity and BERT-based Framework for Unstructured Language Analytics). This name reflects three main components: Named Entity refers to the Named Entity Recognition (NER) module responsible for extracting semantic entities from unstructured product descriptions; BERT-based indicates the use of BERT embeddings to generate deeper contextual representations; and Unstructured Language Analytics emphasizes the framework's focus on processing free-form, non-standard product descriptions, which are the primary source of cold start problems. The choice of the name NEBULA also carries philosophical meaning, as in astronomy, a nebula is known as the "birthplace of stars". This metaphor aligns with the framework's purpose of transforming raw, unstructured data into meaningful knowledge, generating new insights, and producing relevant recommendations even under cold start conditions.

The dataset used in this study consists of 1,000 descriptions of traditional handicraft products collected from various marketplaces. Each product entry includes metadata such as product name, product description, and user ratings. As an illustration, Table II presents five sample product data obtained from the marketplace, which form part of the dataset used in this study.

Table II presents five sample product data obtained from the marketplace, consisting of product name, description, and rating. Product descriptions on marketplaces are generally written freely by sellers without following a standardized format. As a result, the text is considered unstructured, since it varies in writing style, vocabulary, and terminology, often leading to inconsistency. This poses a challenge in the text processing stage, making it necessary to apply Natural Language Processing (NLP) techniques to extract key entities such as product name, material, and origin. Moreover, unstructured descriptions may appear in both Indonesian and English, which requires the system to handle multilingual processing efficiently.

The annotation stage is carried out to label the main entities in the product description. The annotation process uses a Named Entity Recognition (NER) scheme with three types of entities, namely: Product (name or type of craft), Material (main material used), and Origin (region or location of origin of the product). Annotation is done manually by a team of annotators using the BIO (Begin, Inside, Outside) format to ensure consistency in entity tagging. An example of the annotation results can be seen in Fig. 2.

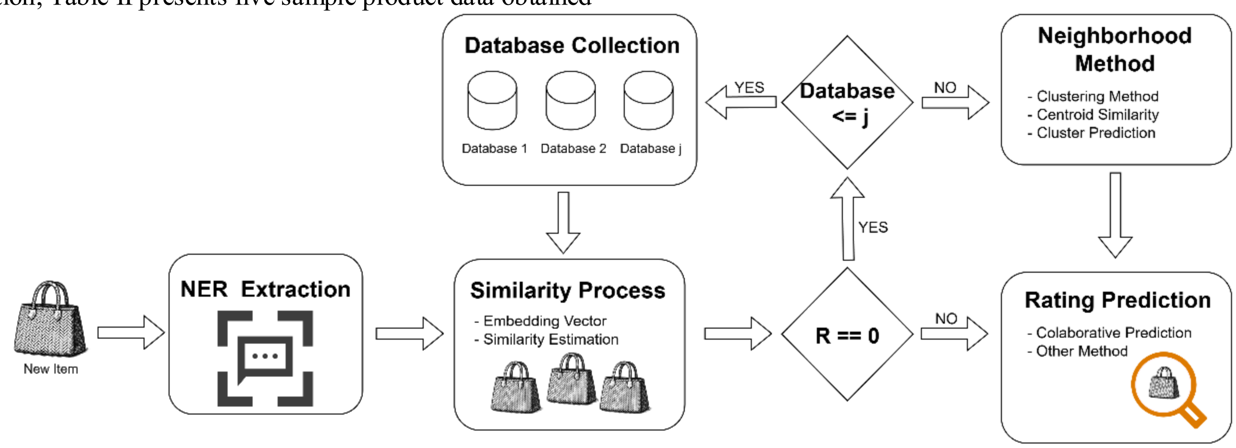

Fig. 1. Framework of the proposed recommendation rating.

TABLE II. SAMPLE DATASET

Product Name	Description	Rating
Ayam Antik Bambu	Ayam Antik, 30 cm, Bambu, merupakan patung dekoratif yang dipahat dari bambu kering dengan teknik ukir alami. Dibuat oleh seniman di Desa Batuan, cocok sebagai pemanis interior yang unik dan bernuansa etnik.	4.2
Gantungan Kunci Barong	Gantungan Kunci Barong ini dibuat dari kayu pule yang dipahat dan dilukis secara manual dengan detail wajah Barong. Produk ini berfungsi sebagai hiasan gantungan kunci yang mencerminkan budaya spiritual Bali. Warna dominan merah, hitam, dan emas memberikan kesan eksotis. Dibuat oleh pengrajin di Desa Mas, Ubud, yang terkenal akan keahlian ukiran kayunya.	4.6
Tali Tasel	Tali Tasel dari Tabanan, Bali, terbuat dari mendong asli yang berkualitas.	4.6
Duck-Shaped Wooden Pull Toy	This Wooden Pull Toy in the shape of a Duck is made from lightweight albasia wood and painted with bright, non-toxic colors. Produced by artisans from Desa Tegallalang, this toy is safe for children and evokes the nostalgia of traditional games.	4.7
Woven Rattan Bracelet	This woven rattan bracelet is made from authentic Balinese rattan, lightweight and comfortable to wear, suitable for enhancing your everyday look.	4.3

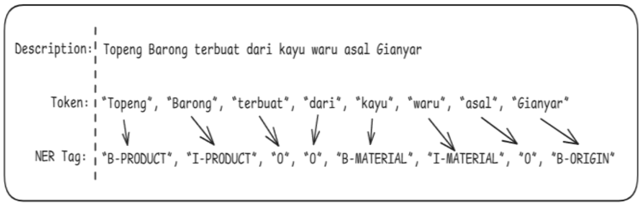

Fig. 2. Data annotation.

Next, the dataset is then split into training and testing sets using an 80:20 ratio, with performance evaluation conducted using precision, recall, and F1-score to measure the accuracy of entity extraction, and Mean Absolute Error (MAE) to evaluate rating prediction errors, while expert validation by two experts in recommendation systems and natural language processing was carried out to assess the relevance of recommendation results for new products.

In the first stage, when a new item N is added to the system, descriptive information about the product is processed using the Named Entity Recognition (NER) technique [25]. This descriptive information is text that contains explanations about the product and is the main source for extraction to determine the entities of the product. Through the NER process, the system identifies and classifies important entities in the text, such as product names, materials, and locations. The identified entities are then compiled into a set $E_N = \{e_1, e_2, \dots, e_k\}$ that represents the characteristics of item N.

Next, the system accesses set C, which has also undergone the previous NER process, generating entity representation E_{C_j} for each item $C_j \in C$. The next stage is the similarity estimation process, in which the system converts the entity set into vector representations using the embedding method. The embedding method is selected adaptively based on data characteristics. For example, BERT (Bidirectional Encoder Representations from Transformers) is used for descriptions with complex contexts, especially in Indonesian, as it can understand complex sentence structures [26]. Meanwhile, spaCy is more suitable for general, lightweight entities, such as descriptions in English, as it is more efficient and faster [27]. Other alternative embedding methods can also be used depending on the needs and type of data being processed.

After that, the similarity score is calculated using the cosine similarity method, which was chosen for its simplicity and ability to effectively measure the closeness between text representations. An item C_j will be included in the reference set $R \subseteq C$ if it has a similarity value with the new item N that exceeds the set threshold, which is 0.70 [28]. Thus, R contains a collection of products in the database that are considered relevant as a reference for predicting the rating of new items.

After the reference set R has been determined, a check will be performed to see whether the set R is empty or not. If the set R has a value, the system will estimate the initial rating $\hat{r}_N$ for the new item N using the weighted average method of the ratings of the items in R, where the weight is given based on the similarity score of each item. However, if the reference set R is an empty set ($R = \emptyset$), the system will trigger the next adaptive mechanism, which is to switch to another database available in the system's database collection. The reanalysis process for items in the new database is performed starting from NER extraction, embedding, to similarity estimation to try to form a new reference set R. If items with sufficient similarity are found in this process, the rating prediction process can be performed immediately.

However, if all databases have been checked and no items with similarity values exceeding the threshold are found ($R = \emptyset$ overall), the system will activate a clustering-based fallback strategy. In this strategy, all existing products in the repository are clustered into several groups using methods such as Probabilistic Neural Network (PNN) or Generalized Lloyd Algorithm (GLA). Each cluster has a centroid vector calculated from the average embedding of the items in that cluster [29]. The new product N is then compared with all cluster centroids using cosine similarity, and the system selects the cluster with the highest similarity. After that, the system calculates the average rating of all products in the selected cluster [30]. This value is then used as the rating prediction for the new product N. The flowchart of the proposed method can be seen in Fig. 3 below.

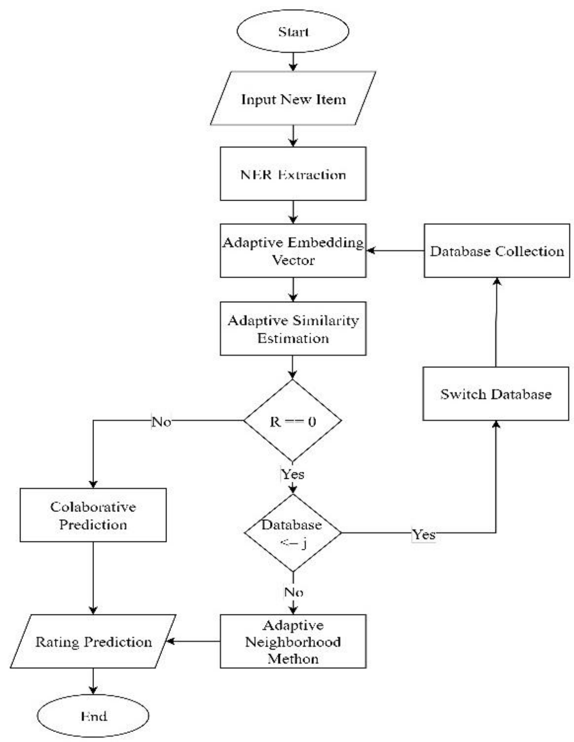

Fig. 3. The proposed model flowchart.

This adaptive mechanism constitutes the novelty of the NEBULA framework. Unlike prior content-based or metadata-driven cold start solutions, NEBULA explicitly integrates multilingual NER (spaCy for English and BERT for Indonesian), dynamic database switching when no similar items surpass the threshold, and a clustering-based fallback strategy to guarantee recommendation continuity. These three layers of adaptability ensure that the system remains robust under varying data conditions, addressing the cold start problem not only algorithmically but also strategically at the data and application levels.

IV. Model Analysis

A. Item Extraction

The first step in the proposed recommendation system is the Item Extraction process, which involves extracting descriptions to search for entities based on the metadata of traditional handicraft products. This process utilises Named Entity Recognition (NER) techniques to extract important entities such as product type (bags, statues, etc.), material type (wood, bamboo, fabric), and product origin location (Gianyar, Tabanan, Karangasem).

Each new product N that is entered will undergo this process to generate entities from the product, in the form of attribute-value pairs derived from the description text (E_N). This representation forms the basis for calculating similarity between products (Algorithm 1).

Algorithm 1: NER Extraction

```
Input: NewProduct
Output: ExtractedEntities
Function ExtractEntities (NewProduct):
    description ← NewProduct.description
    entities ← NER (description)
    return entities
```

B. Item Similarity

After obtaining the entity extraction results in the form of E_N, the system proceeds to the item similarity measurement stage. Each E_N representation generated from the NER process on the product will be converted into an embedding vector using an embedding method that is appropriate for the data characteristics, whether using BERT or spaCy.

Similarity measurement is performed using the cosine similarity method, which calculates the similarity angle between two embedding vectors in a high-dimensional space [13]. This score reflects the extent to which two products share similar meanings or attributes. If v_N is the embedding vector of the new product N and v_{C_j} is the embedding vector of product C_j in the database, then the similarity score is calculated using Eq. (1).

$$\mathrm{sim}\left(N, C_j\right) = \frac{v_N \cdot v_{C_j}}{\|v_N\| \times \left\|v_{C_j}\right\|} \tag{1}$$

The result of this process is a list of similarity scores between the new product and all products in the database. The system then filters products with scores above a certain similarity threshold (0.70) [31], forming a set of reference products R that are considered similar.

However, if the set R is empty (no products in the database have a similarity score exceeding the threshold), the system will automatically move to another available database and repeat the similarity search process using the same procedure. This process continues until all databases have been accessed and compared. If, after searching the entire database, the system still cannot find a product that meets the similarity threshold, it will activate a fallback strategy, which is an alternative mechanism that uses a clustering-based approach to group products with similar metadata characteristics (Algorithm 2).

Algorithm 2: Calculating Similarity Score

```
Input: NewEntities, ListProductDatabases
Output: SimilarProducts or FallbackRating
Function Similarity (NewEntities, ListProductDatabases):
    SimilarProducts ← []
    NewEmbed ← BERT (NewEntities)
    For each Database in ListProductDatabases:
        For each Product in Database:
            ExistingEntities ← Product.Entities
            OldEmbed ← Product.Embedding
            score ← CosineSimilarity (NewEmbed, OldEmbed)
            If score ⩾ Threshold:
                SimilarProducts.append({Product, score})
        If SimilarProducts is not empty:
            return SimilarProducts
    FallbackRating ← FallbackClustering, ListProductDatabases)
    return FallbackRating
```

C. Ratings Prediction

The final stage of the model is predicting ratings for new products. After finding a number of similar reference products, the system calculates the estimated rating based on the weighted average of the ratings of those reference products [32] (Algorithm 3). The higher the similarity of a reference product to the new product, the greater its contribution to the rating estimate. The rating estimate for the new product is calculated using Eq. (2).

$$\hat{r}_N = \frac{\sum_{C_j \in R} \mathrm{sim}(N, C_j) \cdot r_{C_j}}{\sum_{C_j \in R} \mathrm{sim}\,(N, C_j)} \tag{2}$$

Algorithm 3: Calculating Rating Predictions

```
Input: SimilarProductsList, ProductRatings
Output: PredictedRating
Function PredictRating (SimilarProductsList):
    totalWeight ← 0
    weightedSum ← 0
    For each item in SimilarProductsList:
        Product ← item.Product
        Score ← item.SimilarityScore
        Rating ← Product.Rating
        weightedSum ← weightedSum + (Score × Rating)
        totalWeight ← totalWeight + Score
    PredictedRating ← weightedSum / totalWeight
    return PredictedRating
```

V. Performance Assessment

The system was evaluated to measure the effectiveness of the proposed approach in addressing the cold start problem for new items. The assessment was conducted using two approaches: quantitative evaluation (based on classification metrics) and qualitative evaluation (through expert validation).

A. Precision, Recall, and F1-Score

System performance was evaluated using three main metrics, namely precision, recall, and F1-score [33]. These three metrics were used to assess the effectiveness of the Named Entity Recognition (NER) model in extracting descriptive entities from new items. The model was trained for 15 epochs with a precision value of 0.838, recall of 0.846, F1-score of 0.842, and accuracy of 0.967.

These results indicate that the system is capable of extracting important entities (products, ingredients, origin) with high precision and completeness, which directly impacts the quality of embedding formation and the accuracy of similarity estimation between items.

As a comparison, the conventional item-item collaborative filtering method based on cosine similarity and Pearson correlation coefficient, adapted from Hasan et al. [34], was reimplemented using the same dataset as this study. Comparison of evaluation matrices can be seen in Table III.

TABLE III. EVALUATION METRICS FOR THE PROPOSED METHOD AND THE CONVENTIONAL METHOD

Metric	Proposed Method	Conventional Method	Improvement (%)
Precision	0.838	0.725	+15.59%
Recall	0.846	0.720	+17.50%
F1-score	0.842	0.722	+16.54%

Based on the results in Table III, it shows that the proposed method outperforms the conventional method in all metrics, with improvements of 15.59% in precision, 17.50% in recall, and 16.54% in F1-score. The highest improvement in recall demonstrates the system's ability to recognise more relevant entities while maintaining prediction accuracy.

B. Training and Validation

The training and validation processes were conducted over 15 epochs on the Named Entity Recognition (NER) model using a dataset divided into two parts: training data for updating model weights and validation data for measuring generalisation ability.

At the start of training, the training loss began at 0.7015 in the first epoch and decreased significantly to 0.0005 in epochs 14 and 15. This pattern indicates that the model quickly learned patterns from the training data. Meanwhile, the validation loss also decreases significantly from 0.4736 to around 0.1367 at the end of training, indicating improved generalisation without significant signs of overfitting.

Overall, this loss value trend indicates that the training process successfully improved the model's performance significantly while maintaining its generalization ability on new data. This suggests that the choice of training parameters, such as learning rate, number of epochs, and batch size, as well as the model architecture used, are appropriate for learning the patterns of entities present in the dataset. As a result, the trained NER model can be relied upon to accurately extract entities from the available product description data. A more detailed visualization of the training and validation loss trends can be seen in Fig. 4.

Fig. 4. Training loss and validation loss.

C. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) measurement is used to evaluate the accuracy of rating predictions on new items. A lower MAE value indicates predictions that are closer to the actual rating.

A comparison was made between the proposed method, the method of Kadyanan et al. [35] which combines ICHM, Slope One, and SAW, and the method of Liu et al. [36] which applies an adaptive meta-learning approach with an attention mechanism (AMeLU). Testing using the same dataset yielded the following results in Table IV.

TABLE IV. COMPARISON OF MEAN ABSOLUTE ERROR (MAE)

Method	Lowest MAE	Highest MAE
Proposed Method	0.008	0.319
Kadyanan et al. (2019)	0.110	1.060
Liu et al. (2023)	0.75	0.90

Table IV shows a comparison of the Mean Absolute Error (MAE) values between the proposed method, the method of Kadyanan et al. [35], and Liu et al. [36]. The results show that the method of Kadyanan et al. obtained the lowest MAE value of 0.110, while the proposed method achieved 0.008, indicating a significant improvement in prediction accuracy. On the other hand, the highest MAE value for the Kadyanan et al. method reached 1.060, whereas the proposed method only reached 0.319, demonstrating better stability across scenarios. Furthermore, Liu et al. maintained MAE values within the range of 0.75–0.90, which is more stable than Kadyanan et al. but still less accurate compared to the proposed method. These findings confirm that the proposed method is capable of minimising prediction errors more effectively, especially in cold-start scenarios, thereby producing more consistent and accurate estimates than the benchmark methods.

D. Expert Validation

To complement the quantitative evaluation, manual validation (expert validation) was performed by two experts namely Validator 1 and Validator 2 in the field of recommendation systems and natural language processing [37]. This validation aimed to assess the extent to which the system was able to generate descriptions of new products that matched products with a history of user interaction. The validation process was carried out by comparing the system's similarity results with the assessments of the two validators, as shown in Table V [38].

TABLE V. VALIDATOR ASSESSMENT CONTINGENCY

	Validator 2 (Valid)	Validator 2 (Invalid)	Total
Validator 1 (Valid)	910	30	940
Validator 1 (Invalid)	10	50	60
Total	920	80	1000

Table V shows the results of the contingency assessment validator, which compares the assessments of two validators on the output of the recommendation system. Symbol a represents the number of data points deemed valid by both validators, totaling 910 data points. b refers to data points deemed valid by Validator 1 but invalid by Validator 2, totaling 30 data points. c denotes data points deemed invalid by Validator 1 but valid by Validator 2, totaling 10 data points. while d represents data deemed invalid by both validators, totaling 50 data points. Thus, the total number of validated data points is T = 1000. Consequently, the agreement between the two validators is represented by the values of a and d, while the disagreement is indicated by the values of b and c. To measure the level of agreement among validators, Cohen's Kappa is used as shown in Eq. (3), (4), and (5).

Actual agreement proportion (P_O):

$$P_O = \frac{a+d}{T} \tag{3}$$

Proportion of chance agreement (P_e):

$$P_e = \left(\frac{a+b}{T} \cdot \frac{a+c}{T}\right) + \left(\frac{c+d}{T} \cdot \frac{b+d}{T}\right) \tag{4}$$

Kappa (K):

$$K = \frac{P_O - P_e}{1 - P_e} \tag{5}$$

Based on the data in Table III, $P_O = 0.96$, $P_e = 0.869$, and K = 0.69 were obtained. According to the interpretation by Madadizadeh et al. [39], Kappa values can be categorized as follows: < 0.00 = Poor agreement, 0.00–0.20 = Slight agreement, 0.21–0.40 = Fair agreement, 0.41–0.60 = Moderate agreement, 0.61–0.80 = Substantial agreement, and 0.81–1.00 = Almost perfect agreement. Thus, a Kappa value of 0.69 falls into the substantial agreement category, indicating a high level of agreement between the two validators. This result confirms that the recommendations generated by the system are deemed consistent and relevant by the experts, thereby strengthening the model's reliability in addressing the cold start problem.

VI. CONCLUSION

This study proposes a metadata-based recommendation system model using a Natural Language Processing (NLP) approach, specifically Named Entity Recognition (NER), to address the cold start problem, particularly in traditional handicraft products. The system extracts descriptive entities such as product type, materials, and regional origin from unstructured text, then converts them into vector representations using adaptive embedding methods. The model integrates a similarity estimation mechanism with a clustering-based fallback strategy to ensure rating predictions can still be made even if no similar items are found in the database.

Experimental results show that the NER model achieves excellent performance with precision of 0.838, recall of 0.846, F1-score of 0.842, and accuracy of 0.967, reflecting the system's ability to accurately extract important entities. Compared to conventional filtering approaches, this model demonstrates significant improvements across all evaluation metrics. Additionally, the system's prediction performance was measured using Mean Absolute Error (MAE), with an average of 0.159, a minimum value of 0.008, and a maximum value of 0.319. The relatively low MAE value indicates that the system's rating estimates are very close to the actual values, thereby reinforcing the model's reliability in the context of recommendations. Validation conducted by two experts showed a substantial agreement level, indicating that the recommendations generated by the system are considered consistent and relevant, especially when dealing with new items (cold start).

Scientifically, this research proves that integrating metadata with NER can be an effective alternative in overcoming the cold start problem, while expanding the application of content-based methods in the local cultural domain. From a practical perspective, this model has the potential to support SMEs and local marketplaces in enhancing the visibility of craft products and promoting cultural initiatives in the digital realm.

This study still has several limitations that can be explored in future research. The recommendation system can be extended by integrating both explicit and implicit user feedback, enabling it to address the cold start problem not only at the item level but also for new users. In addition, the results of Named Entity Recognition (NER) can be enriched with semantic approaches, such as knowledge graphs or semantic embeddings, to better capture contextual meaning and relationships among entities. The proposed framework also has potential to be applied across other product domains, such as fashion, culinary, and tourism, to evaluate scalability and ensure robustness in broader and multilingual contexts. Furthermore, the fallback clustering mechanism can be further developed to handle cases where products have no similarity values above the defined threshold with items in the database. In such scenarios, the system can still provide relevant recommendations by employing alternative clustering approaches, such as Graph-based Latent Allocation (GLA), Incremental Clustering using Hierarchical Merging (ICHM), K-Means Variants, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), or Self-Organizing Maps (SOM).

ACKNOWLEDGMENT

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] L. Mohanty, L. Saraswat, P. Garg, and S. Lamba, “Recommender Systems in E-Commerce,” Proc. - 2022 5th Int. Conf. Comput. Intell. Commun. Technol. CCICT 2022, no. July, pp. 114–119, 2022, doi: 10.1109/CCiCT56684.2022.00032.

[2] Vrushal Jagtap, Om Mahajan, Bhavesh Rane, Kumar Kolhe, and Priyanka Bendale, “General Purpose Recommendation System,” Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 1, pp. 278–280, 2023, doi: 10.48175/ijarsct-9040.

[3] R. S. Sharma, A. A. Shaikh, and E. Li, “Designing Recommendation or Suggestion Systems: looking to the future,” Electron. Mark., vol. 31, no. 2, pp. 243–252, 2021, doi: 10.1007/s12525-021-00478-z.

[4] M. Ibrahim, I. S. Bajwa, N. Sarwar, F. Hajjej, and H. A. Sakr, “An Intelligent Hybrid Neural Collaborative Filtering Approach for True Recommendations,” IEEE Access, vol. 11, no. June, pp. 64831–64849, 2023, doi: 10.1109/ACCESS.2023.3289751.

[5] R. Khatwal and S. N. Hasan, “Cold Start Problem in Recommendation System: A Solution Model Based on Clustering and Association Rule Techniques,” 2022 5th Int. Conf. Multimedia, Signal Process. Commun. Technol. IMPACT 2022, pp. 1–8, 2022, doi: 10.1109/IMPACT55510.2022.10029293.

[6] N. AlRossais, D. Kudenko, and T. Yuan, Improving cold-start recommendations using item-based stereotypes, vol. 31, no. 5. Springer Netherlands, 2021. doi: 10.1007/s11257-021-09293-9.

[7] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, “Collaborative filtering and deep learning based recommendation system for cold start items,” Expert Syst. Appl., vol. 69, pp. 1339–1351, 2017, doi: 10.1016/j.eswa.2016.09.040.

[8] A. Gharahighehi, K. Pliakos, and C. Vens, “Addressing the Cold-Start Problem in Collaborative Filtering Through Positive-Unlabeled Learning and Multi-Target Prediction,” IEEE Access, vol. 10, no. November, pp. 117189–117198, 2022, doi: 10.1109/ACCESS.2022.3219071.

[9] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, “Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering,” J. Appl. Data Sci., vol. 4, no. 3, pp. 289–302, 2023, doi: 10.47738/jads.v4i3.115.

[10] W. Shalaby, S. Oh, A. Afsharinejad, S. Kumar, and X. Cui, M2TRec: Metadata-aware Multi-task Transformer for Large-scale and Cold-start free Session-based Recommendations, vol. 1, no. 1. Association for Computing Machinery, 2022. doi: 10.1145/3523227.3551477.

[11] K. N. Mishra, A. Mishra, P. N. Barwal, and R. K. Lal, “Natural Language Processing and Machine Learning-Based Solution of Cold Start Problem Using Collaborative Filtering Approach,” Electron., vol. 13, no. 21, 2024, doi: 10.3390/electronics13214331.

[12] J. Barroca, A. Shivkumar, B. Q. Ferreira, E. Sherkhonov, and J. Faria, “Enriching a Fashion Knowledge Graph from Product Textual Descriptions,” CEUR Workshop Proc., vol. 3184, pp. 36–51, 2022.

[13] J. Zhang, F. Wang, F. Ma, and G. Song, “Text Similarity Calculation Method Based on Optimized Cosine Distance,” Proc. - 2022 Int. Conf. Electron. Devices, Comput. Sci. ICEDCS 2022, pp. 37–39, 2022, doi: 10.1109/ICEDCS57360.2022.00015.

[14] C. Xiao et al., “UPRec: User-aware Pre-training for sequential Recommendation,” AI Open, vol. 4, no. September, pp. 137–144, 2023, doi: 10.1016/j.aiopen.2023.08.008.

[15] H. Zhou, F. Xiong, and H. Chen, “A Comprehensive Survey of Recommender Systems Based on Deep Learning,” Appl. Sci., vol. 13, no. 20, 2023, doi: 10.3390/app132011378.

[16] L. Jiang, L. Liu, J. Yao, and L. Shi, “A hybrid recommendation model in social media based on deep emotion analysis and multi-source view fusion,” J. Cloud Comput., vol. 9, no. 1, 2020, doi: 10.1186/s13677-020-00199-2.

[17] E. S. Negara, Sulaiman, R. Andryani, P. H. Saksono, and Y. Widyanti, “Recommendation System with Content-Based Filtering in NFT Marketplace,” J. Adv. Inf. Technol., vol. 14, no. 3, pp. 518–522, 2023, doi: 10.12720/jait.14.3.518-522.

[18] S. Lestari, Yulmaini, Aswin, S. Y. Ma’ruf, Sulyono, and R. R. N. Fikri, “Alleviating cold start and sparsity problems in the micro, small, and medium enterprises marketplace using clustering and imputation techniques,” Int. J. Electr. Comput. Eng., vol. 14, no. 3, pp. 3220–3229, 2024, doi: 10.11591/ijece.v14i3.pp3220-3229.

[19] A. Panteli and B. Boutsinas, “Addressing the Cold-Start Problem in Recommender Systems Based on Frequent Patterns,” Algorithms, vol. 16, no. 4, 2023, doi: 10.3390/a16040182.

[20] P. Gupta, A. Deo, S. S. Khan, A. S. Rajput, K. Joshi, and A. Chourasia, “A Progressive Approach for a Personalized Recommendation System Using Hybrid Deep Learning,” Ing. des Syst. d’Information, vol. 29, no. 5, pp. 1999–2009, 2024, doi: 10.18280/isi.290531.

[21] M. T. Ramakrishna et al., “HCoF: Hybrid Collaborative Filtering Using Social and Semantic Suggestions for Friend Recommendation,” Electron., vol. 12, no. 6, pp. 1–23, 2023, doi: 10.3390/electronics12061365.

[22] Y. Afoudi, M. Lazaar, and M. Al Achhab, “Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network,” Simul. Model. Pract. Theory, vol. 113, no. April, p. 102375, 2021, doi: 10.1016/j.simpat.2021.102375.

[23] A. Vomberg, N. Schauerte, S. Krakowski, C. Ingram Bogusz, M. J. Gijsenberg, and A. Bleier, “The cold-start problem in nascent AI strategy: Kickstarting data network effects,” J. Bus. Res., vol. 168, no. December 2022, p. 114236, 2023, doi: 10.1016/j.jbusres.2023.114236.

[24] Z. P. Zhang, Y. Kudo, T. Murai, and Y. G. Ren, “Addressing complete new item cold-start recommendation: A niche item-based collaborative filtering via interrelationship mining,” Appl. Sci., vol. 9, no. 9, 2019, doi: 10.3390/app9091894.

[25] X. Shen, X. Hu, N. Liu, and Y. Shen, “Named Entity Recognition of Multilevel Features,” 2022 4th Int. Conf. Front. Technol. Inf. Comput. ICFTIC 2022, pp. 575–578, 2022, doi: 10.1109/ICFTIC57696.2022.10075239.

[26] P. Chaudhry, “Bidirectional Encoder Representations from Transformers for Modelling Stock Prices,” Int. J. Res. Appl. Sci. Eng. Technol., vol. 10, no. 2, pp. 896–901, 2022, doi: 10.22214/ijraset.2022.40406.

[27] B. B. Elov, “Making Use of a ‘Spacy’ Module in the Natural Language Processing,” J. Sci. Innov. Dev., vol. 5, no. 4, pp. 41–54, 2022, doi: 10.36522/2181-9637-2022-4-5.

[28] A. Senior, “Pearson ’ s and Cosine Correlation,” Int. Conf. Trends Electron. Informatics ICEI 2017 Calc., vol. 05 June 20, no. May, pp. 1000–1004, 2017.

[29] M. Ghiasabadi Farahani, J. Akbari Torkestani, and M. Rahmani, “Adaptive personalized recommender system using learning automata and items clustering,” Inf. Syst., vol. 106, p. 101978, 2022, doi: 10.1016/j.is.2021.101978.

[30] M. A. Al-Hagery, “A novel Based-Approach Composed of Clustering Algorithm & Cosine Similarity for Products Recommendation,” Int. J. Educ. Inf. Technol., vol. 14, pp. 133–141, 2020, doi: 10.46300/9109.2020.14.16.

[31] M. D. Khilji, “Features Matching using Natural Language Processing,” Int. J. Cybern. Informatics, vol. 12, no. 2, pp. 251–260, 2023, doi: 10.5121/ijci.2023.120218.

[32] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, “Facing the cold start problem in recommender systems,” Expert Syst. Appl., vol. 41, no. 4 PART 2, pp. 2065–2073, 2014, doi: 10.1016/j.eswa.2013.09.005.

[33] S. Al Rawahi, N. Al Musalhi, and E. Celebi, “Recommendation Systemfor Daily Consumer Purchases List Usingspecial Hybrid Approach,” vol. 30, no. 4, pp. 1239–1250, 2024, doi: 10.53555/kuey.v30i4.1642.

[34] M. Hasan and F. Roy, “An item–item collaborative filtering recommender system using trust and genre to address the cold-start problem,” Big Data Cogn. Comput., vol. 3, no. 3, pp. 1–15, 2019, doi: 10.3390/bdcc3030039.

[35] I. G. A. G. A. Kadyanan, I. B. G. Dwidasmara, I. B. M. Mahendra, I. K. Ari Mogi, and I. W. Puguh Sudarma, “The design of typical Balinese food recommendation system using hybrid method of collaborative filtering and Slope One algorithm on mobile device platform,” ACM Int. Conf. Proceeding Ser., pp. 111–116, 2019.

[36] S. Liu, Y. Liu, X. Zhang, C. Xu, J. He, and Y. Qi, “Improving the Performance of Cold-Start Recommendation by Fusion of Attention

Network and Meta-Learning," Electron., vol. 12, no. 2, pp. 1–14, 2023, doi: 10.3390/electronics12020376.

[37] W. Al Jurdi, J. B. Abdo, J. Demerjian, and A. Makhoul, "Group Validation in Recommender Systems: Framework for Multi-layer Performance Evaluation," ACM Trans. Recomm. Syst., vol. 2, no. 1, pp. 1–25, 2024, doi: 10.1145/3640820.

[38] K. S. Tan, Y. C. Yeh, P. S. Adusumilli, and W. D. Travis, "Quantifying Interrater Agreement and Reliability Between Thoracic Pathologists: Paradoxical Behavior of Cohen's Kappa in the Presence of a High Prevalence of the Histopathologic Feature in Lung Cancer," JTO Clin. Res. Reports, vol. 5, no. 1, p. 100618, 2024, doi: 10.1016/j.jtocrr.2023.100618.

[39] F. Madadizadeh, H. Ghafari, and S. Bahariniya, "The Open Public Health Journal Kappa Statistics: A Method of Measuring Agreement in Dental Examinations," Open Public Health J., vol. 16, pp. 1–6, 2023, doi: 10.2174/0118749445259818231016040344.