

RESEARCH ARTICLE

NLP-Based Recommendation Approach for Diverse Service Generation

BAEK JEONG^{ID} AND KYOUNG JUN LEE

Department of Big Data Analytics, Kyung Hee University, Seoul 02447, South Korea

Corresponding author: Kyoung Jun Lee (klee@khu.ac.kr)

This work was supported in part by the Brain Korea 21 (BK21) Fostering Outstanding Universities for Research (FOUR) funded by the Ministry of Education (MOE), South Korea; and in part by the National Research Foundation of Korea (NRF).

ABSTRACT In this study, we examine the potential of language models for natural language processing (NLP)-based recommendations, with a distinct focus on predicting users' next product purchases based on their prior purchasing patterns. Our model specifically harnesses tokenized rather than complete product names for learning. This granularity allows for a refined understanding of the interrelations among different products. For instance, items like 'Chocolate Milk' and 'Coffee Milk' find linkage through the shared token 'Milk.' Additionally, we explored the impact of various n-grams (unigrams, bigrams, and trigrams) in tokenization to further refine our understanding of product relationships and recommendation efficacy. This nuanced method paves the way for generating product names that might not exist in current retail settings, exemplified by concoctions like 'Coffee Chocolate Milk.' Such potential offerings can provide retailers with fresh product brainstorming opportunities. Furthermore, scrutiny of the frequency of these generated product name tokens can reveal prospective trends in purchasing keywords. This facilitates enterprises in creative brainstorming of novel products and swiftly responding to the dynamic demands and trends of consumers. The datasets used in this study come from UK e-Commerce and Instacart Data, comprising 71,205 and 166,440 rows, respectively. This investigation juxtaposes the NLP-based recommendation model, which employs tokenization, with its non-tokenized counterpart, leveraging Hit-Rate and mean reciprocal rank (MRR) as evaluative benchmarks. The outcomes distinctly favor the tokenized NLP-based recommendation model across all evaluated metrics.

INDEX TERMS Natural language processing, recommendation, tokenization, n-grams, service generation, new product brainstorming, keyword trend forecasting.

I. INTRODUCTION

Recently, there has been significant research on natural language processing (NLP) in the field of artificial intelligence, leading to rapid advancements in text processing technology. A significant factor contributing to this growth is the latest advancements in deep learning, which have elevated NLP performance to unprecedented levels, drawing considerable attention from the academia and industry [1], [2]. The increasing focus on NLP research can be attributed to several factors, but the emergence of innovative algorithms and models, such as the Transformer, BERT, and GPT, has played a pivotal role [3], [4]. Advancements in NLP technology, which

enables machines to understand, generate, and communicate text, offer new opportunities across various industry sectors. Such NLP technologies offer businesses critical assistance in enhancing information processing efficiency, improving customer service, and devising new marketing strategies. Given this potential, there has been an intensified interest in NLP research within the industry, with both researchers and enterprises striving for further refinement and expansion of NLP capabilities.

The development of recommendation systems is increasingly recognized as a crucial application area for NLP [5]. Text sequences, such as news or articles, can provide essential information for recommending highly relevant content to users through the analysis of meaning or themes using NLP. Furthermore, by analyzing user reviews and feedback,

The associate editor coordinating the review of this manuscript and approving it for publication was Anandakumar Haldorai^{ID}.

one can discern user preferences and reactions, and integrating this information into a recommendation algorithm enables personalized content recommendation [6]. Notably, a crucial intersection between NLP and recommendation systems is the utilization of sequence information [7]. Sequence data, such as user behavior patterns, purchase history, and search records, accumulate over time, allowing for a precise reflection of a user’s recent actions and evolving preferences. NLP research deeply explores the context of text, allowing for a precise understanding of the meanings of sentences or documents derived from sequential patterns. Sequence modeling techniques have been developed to accurately predict specific or subsequent words within a given context.

The aim of this study was to extend the application of text sequence processing techniques beyond merely predicting the next word to forecasting what product will be purchased next. Specifically, based on a user’s purchase history, we develop a methodology that allows recommendation systems to accurately predict and suggest the product that the user is most likely to purchase next. Traditional research on recommendation systems incorporating NLP has focused on sequential recommendation, characterizing recommendations derived from product-level learning [5], [7], [8], [9]. This has led to a significant oversight: the nuanced connections shared by products with common elements are often overlooked when analysis is conducted without tokenizing product names. For example, ‘Chocolate Milk’ and ‘Strawberry Milk’ are treated as distinct entities, although they share a common ‘Milk’ component. This limitation in the current body of research underlines the need for a more sophisticated approach that we address in this study.

To fully harness the advantages of NLP research, we posit that analyzing product names at the token level will facilitate a more detailed learning of product sequences. By tokenizing product names, such as ‘Chocolate’ ‘Milk’ and ‘Strawberry’ ‘Milk’, they can be linked through the common token ‘Milk’, enabling the system to recognize a broader range of associations and relations. This approach of interlinking various product name tokens and learning from their interactions is illustrated in Figure 1 and serves as the core motivation of our study.

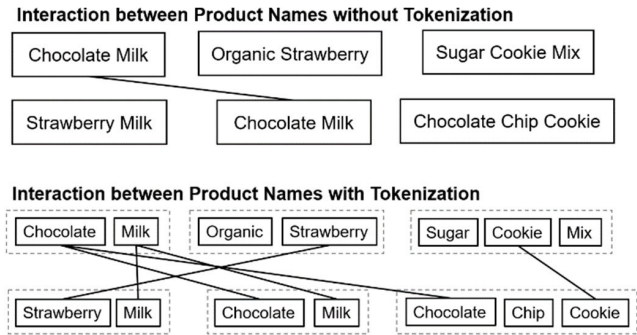


FIGURE 1. Interaction between product names with and without tokenization.

Tokenization in NLP is the critical process of segmenting text into its constituent elements, or tokens, which can be as small as words or as significant as phrases. Each token represents a fundamental unit of meaning, crucial for the machine’s understanding and analysis of language. This step is especially challenging given the diverse and complex grammatical structures present across languages [10]. The methodology chosen for tokenization directly impacts the performance of NLP tasks, with rule-based methods often applied to straightforward texts and more advanced statistical or neural network approaches employed for intricate linguistic patterns. Subword tokenization is particularly adept at discerning meaningful components within words, an aspect essential for processing languages with rich morphology [11]. In the context of recommendation systems, traditional research has frequently neglected the intricate semantic connections between products, primarily due to non-tokenized approaches. This oversight has led to a practical impact where systems fail to recognize the shared attributes of products, limiting the personalization and relevance of recommendations. To address this gap, we employ N-gram techniques—specifically unigrams, bigrams, and trigrams—as these methods allow for different granularities in recognizing patterns within product names. The unigram model considers individual tokens, while bigrams and trigrams account for pairs and triplets of tokens, providing insight into immediate sequential relationships [12]. Through comparative analysis of these techniques, we aim to determine which yields the most accurate reflection of consumer behavior and, consequently, more precise recommendations. By dissecting product names into tokens, we anticipate uncovering a deeper layer of consumer preference that enhances the personalization of the recommendation process. We aim to explore which tokenization method yields better performance in product name processing. By adopting a token-level approach in product name learning, we expect to achieve a more granular understanding that could lead to more relevant recommendations. However, it is acknowledged that token-based modeling may occasionally generate non-existent product names. For instance, if the ‘Chocolate’ token is followed by ‘Milk’ or ‘Chip’, it aligns with existing products, but a combination like ‘Chocolate Strawberry Milk’ represents a novel, albeit non-existent, product name. This mirrors the hallucination issue commonly found in large language models (LLM) like GPT [13]. While these may seem like errors, they also represent creative opportunities for product development and brainstorming. Furthermore, the ability to predict product trends using NLP-based recommendation systems is a significant advancement. Since the primary objective of recommendation systems is to forecast a user’s next purchase, these novel product names can be viewed as predictive data, offering insights into future trends. Treating tokenized product names as keywords and analyzing their frequency allows us to propose new services based on predicted keyword trends.

Hence, this study aims to enhance recommendation performance by focusing on the tokenization of product names

and exploring diverse service generation strategies. Our primary contributions are substantial: we have created a novel NLP-based recommendation model that operates at the token level of product names. Additionally, we are conducting experiments to further improve performance by comparing token levels using the n-grams approach. Utilizing this model, we have introduced two innovative services: new product brainstorming and keyword trend forecasting. These contributions demonstrate our commitment to advancing the field of NLP and its applications in recommendation systems, overcoming the limitations of previous methodologies, and setting new benchmarks for personalized, predictive commerce.

II. RELATED WORK

To facilitate the development of NLP-based recommendation and its utilization in diverse service generation, we briefly review several pertinent studies.

A. NATURAL LANGUAGE PROCESSING

The field of NLP has seen significant advancements over the past several decades. Initially dominated by rule-based and statistical methods, a major shift occurred with the advent of deep learning techniques. In particular, the introduction of the Transformer architecture has greatly altered the paradigm of the NLP field. The Transformer was first introduced in a paper titled “Attention Is All You Need” in 2017 [14]. The traditional recurrent neural network (RNN) sequentially processes the elements of an input sequence, while long short-term memory (LSTM) is a variant of RNN designed to address long-term dependency issues. However, unlike these traditional RNN and LSTM architectures, the Transformer introduced the self-attention mechanism, allowing the processing of the entire sentence simultaneously without sequential information processing. This enables a more effective understanding of the relationships among words located at various positions within the text. The structure of the Transformer has had a significant impact on two major areas of NLP: natural language understanding (NLU) and natural language generation (NLG) [15], [16], [17]. NLU focuses on the process of computers understanding and interpreting human language; recognizing entities, relationships, and intents within sentences; understanding context; and analyzing meaning. Advanced models based on encoding layers are primarily used in this process, with the BERT model being exemplary. BERT has demonstrated outstanding performance in various NLP tasks by leveraging the advantages of the Transformer architecture to the fullest extent [18]. Meanwhile, NLG focuses on the process of computers generating human-readable natural language text. This involves utilizing machine learning algorithms, template-based systems, etc. to convert data or information into natural text. Advanced models based on decoding layers are primarily utilized in this field, with the GPT model being exemplary. GPT, which is also based on the Transformer architecture, exhibits excellent performance in text generation [3], [19].

B. NLP-BASED RECOMMENDATION

The evolution of recommendation systems has seen a shift from basic collaborative filtering and cold start solution strategies [20], [21], [22], [23] to utilizing complex semantic patterns in text data for fine-grained recommendations [5], [8], [9], [24]. BERT4Rec and similar architectures have shown promise in capturing dynamic user preferences through the analysis of historical behavior. These models, informed by the Transformer’s strengths, outperform traditional sequential neural networks [8]. The incorporation of semi-automatic annotation for sentiment analysis [25] and adapted feature selection algorithms [26] further improves the system’s ability to provide personalized, accurate recommendations. This approach aims to predict future user purchases by learning significant product name characteristics and patterns. These advancements, combined with the full strengths of the Transformer architecture, allow for a deeper understanding and prediction of future user purchases based on semantic information. This paper proposes a new approach, utilizing the full strengths of the Transformer architecture to learn significant characteristics and patterns of product names, aiming to predict future user purchases based on semantic information. HybridBERT4Rec utilizes BERT to extract the characteristics of user interactions with purchased items and provide recommendations based on this. This model sequences users’ historical interactions and is designed to better reflect users’ changing interests [24]. However, most of these studies treated the entire product name as a single unit without tokenization. Although there have been attempts to learn product names as tokens [27], they have the complex limitation of a dual structure that combines a the Transformer and Word2Vec. In this paper, we propose a new approach to recommendation systems, fully leveraging the strengths of the entire Transformer architecture. We explored a method for learning significant characteristics and patterns of product names using an encoder and generating new product names at the token level using a decoder. This approach aims to learn the semantic information of product names related to users’ past purchase patterns and, based on this, predict the product names that users are likely to purchase in the future, which is expected to be greatly beneficial.

C. SERVICE GENERATION

Service generation is the process of designing or innovating new services. Although this approach can vary depending on the industry, field, or company, the overarching goal is to develop services that meet or exceed user expectations [28], [29]. In relation to this, NLP and recommendation systems have been established as key elements in enhancing user experience and innovating service delivery. Specifically, text data generated by users across various digital platforms contain valuable information regarding user preferences, behaviors, and expectations. Companies that deeply understand the significance of these data are exploring strategies for offering personalized services by integrating NLP and

recommendation systems. This approach goes beyond merely analyzing user text data to provide tailored services, playing a crucial role in enhancing service quality, efficiency, and diversity.

Services leveraging NLP have contributed significantly to the provision of more precise and personalized services by utilizing users' text data. Various studies have explored the possibilities and effects of service generation, showcasing the potential in this domain [30], [31], [32], [33]. Traditional research has focused on analyzing text data such as user reviews, feedback, and interactions to understand user preferences and behavior patterns, primarily to provide personalized recommendations. In addition, active research is being conducted to measure user similarities and content-based recommendations using user text data. However, service providers and platforms must customize their models for each specific task. For efficient model development, a strategy for generating diverse services using a single model is required. For instance, large language models like GPT provide various services, including sentence summarization, translation, and document generation [19], [34].

Prior research has highlighted several strategies for service generation using recommendation systems. First, personalized marketing campaigns using recommendation systems delve deeply into users' past purchase histories and search patterns to provide personalized marketing messages or discount coupons, combining users' past purchase histories and product preferences to offer personalized marketing campaigns [35]. Target marketing is the duality of product recommendations targeting users who are expected to purchase certain products [36]. Second, there has been significant interest in product bundle suggestions based on users' purchase histories. For example, after a user purchases a digital camera, a recommendation system can suggest related products, such as memory cards or camera cases [37], [38]. Moreover, recommendation systems can offer bundling of products that appear together as a result [39]. Our study extends the insights of these earlier studies and introduces innovative service generation methods centered on NLP, aiming to explore new service generation strategies by maximizing the advantages of NLP-based recommendation systems.

III. NLP-BASED RECOMMENDATION

In this section, we propose an NLP-based recommendation that utilizes the Transformer to learn product names at the token level, and we validate its performance using a dataset where product name verification is possible.

A. FRAMEWORK

A distinctive feature of this study is the tokenization of product names into individual tokens for the recommendation system, and the framework for the NLP-based recommendation is depicted in Figure 2. Initially, when a list of product names purchased by a user is input, the names are tokenized based on the space within them. For instance, a purchased product name 'Chocolate Milk' is tokenized into 'Chocolate' and 'Milk' based on the space. These tokenized product names are then trained as token sequences using the Transformer, which subsequently predicts the product names to purchase in tokens. As an example, when product names like 'Chocolate', 'Milk', 'Fruit', 'Snack', 'Chocolate', 'Chip', 'Cookie' are input, the Transformer's training results in predictions such as 'Strawberry' and 'Milk'. These tokenized predictions are then detokenized using spaces, converting 'Strawberry' and 'Milk' back into 'Strawberry Milk'. During the training and prediction of product names in tokens, it is possible to generate product names that are not actually sold. As shown in Figure 3, using the Transformer could lead to the generation of a product name like 'Coffee Chocolate Milk', which is not available in actual stores. Because this impedes the recommendation system's ability to suggest actual products, we check against a list of product names to verify the existence of the product and replace non-existent product names using similarity to find the most similar existing product name. To ensure the efficacy of our recommendation system in suggesting actual products, we chose the Jaccard similarity metric for its simplicity and computational efficiency. This contrasts with vector space models that require extensive computation; Jaccard similarity directly compares sets by quantifying the overlap between tokenized product names, which is particularly suitable for our use case. The complexity of product names varies greatly, and our focus is on the presence or absence of shared tokens rather than their frequency or order. The Jaccard similarity, calculated as the proportion of shared tokens to the total unique tokens in both product names, allows for accurate suggestions of existing product names when non-existent ones are generated.

The formula used for Jaccard similarity is as follows:

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \quad (1)$$

The output from the Transformer selects the token with the highest softmax probability as the first in the sequence and subsequently generates the next tokens in an autoregressive manner to form the product name. For instance, if the Transformer's decoder first predicts 'Chocolate', it will then predict the 'milk' token to complete the product name. The subsequent prediction, excluding the highest

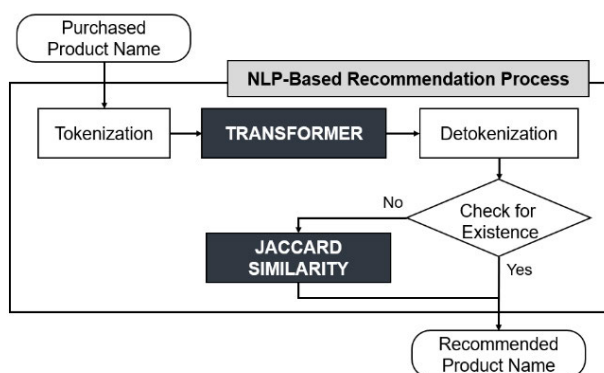


FIGURE 2. Framework of NLP-based recommendation.

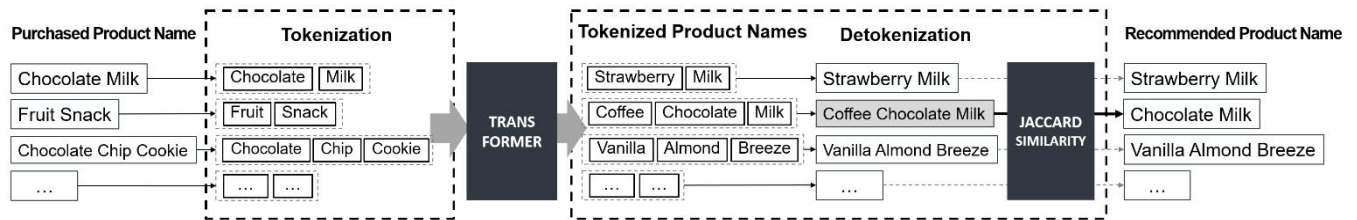


FIGURE 3. NLP-based recommendation process using product names with tokenization.

probability 'Chocolate', will select the next highest probable token 'Vanilla', followed by 'Almond' and 'Breeze', thus forming a product name with tokens 'Vanilla', 'Almond', and 'Breeze', predicting the top- k product names in this manner. Through these steps, the final k recommended product names, trained and predicted as tokens, are derived.

B. DATASETS

For our NLP-based recommendation experiments, we employed two datasets with distinguishable product names:

- **UK e-Commerce¹**: Sourced from a UK-based online retail platform. This dataset aggregates product names and purchase dates documented in user invoices. The platform predominantly features food items, daily essentials, and electronic appliances.
- **Instacart²**: This dataset contains the transactional records of grocery orders, noting the specific week, time, and individual products ordered in the US.

A detailed breakdown of these raw datasets is presented in Table 1.

TABLE 1. Summary of the raw datasets.

Dataset	#Transactions	#Users	#Products
UK e-Commerce	210,367	2,920	3,741
Instacart	33,819,106	206,209	49,688

After examining the raw data, pre-processing was performed. First, we removed errors, such as missing values, from the dataset and chronologically listed the product names purchased by each user. Then, as shown in Figure 4, we grouped the product names into sets of five to form a row, using four product names for training and one product name as the label for each row. To address the issue of repeated product names in our expansive datasets, we utilized a data cleansing technique based on a previous work [9]. To obtain a more comprehensive dataset, we curated an additional collection of products by transferring two pairs of products from every five transactions.

Owing to computational constraints and to maintain coherence in data volume when juxtaposed with the UK e-commerce dataset, we opted to use only 1% of the Instacart

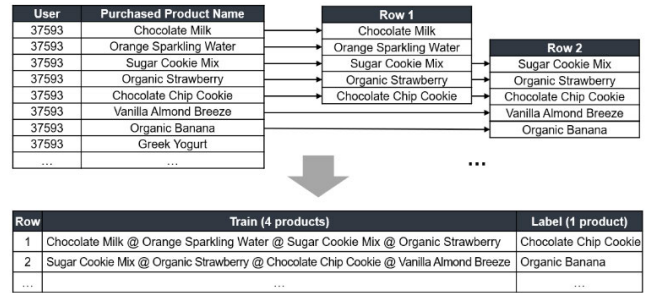


FIGURE 4. Example of dataset preprocessing.

dataset. The descriptions of the two datasets processed at the row level, as previously mentioned, are provided in Table 2. This study employed NLP-based recommendation, tokenizing product names for use. Four product names were allocated to Train and one to Label, and their average tokens are listed in the table. In our experiment, we allocated 80% of the rows for training, 10% for validation, and the remaining 10% for testing.

TABLE 2. Summary of the datasets after preprocessing.

Dataset	#Rows	Avg. #Tokens per Row		
		Total	Train	Label
UK e-Commerce	71,205	22.5	18.0	4.5
Instacart	166,440	19.2	15.4	3.8

C. EVALUATION METRICS

To evaluate the proposed model, we drew inspiration from established studies on recommendation system evaluation metrics [40], [41], [42]. These studies validated our choice of metrics and highlighted their significance in the current context. The following two metrics were employed to assess performance:

- **Hit-Rate**: This metric is prevalent in recommendation systems. It gauges whether the top- K product names suggested to each user are aligned with the product name of their most recent purchase. A match within the top- K recommendations is considered a hit. The Hit-Rate is the ratio of users with hits to the total number of users. The corresponding formula is as follows:

$$\text{Hit-Rate} = \frac{\# \text{Hit Users}}{\# \text{Users}} \quad (2)$$

¹<https://www.kaggle.com/datasets/theopenk/2017online-retail>

²<https://www.kaggle.com/c/instacart-market-basket-analysis>

- **Mean Reciprocal Rank (MRR):** MRR quantifies the ranking of the last purchased product name within the top-K recommended list. A superior MRR indicates the success of the system in recommending relevant product names at the top positions. The formula for MRR is as follows:

$$MRR_k = \frac{1}{K} \sum_{i=1}^K \frac{1}{rank_i} \quad (3)$$

For a comprehensive evaluation, we present our results for both Hit-Rate and MRR at multiple K-values, specifically, $K = 3, 5, 10, 15$, and 20 . This choice helps assess the robustness of the model across various recommendation list lengths. Higher metric values signify enhanced performance.

D. COMPARISON MODEL AND PARAMETER SETTINGS

To verify the superiority of NLP-based recommendation using product names at the token level, we set up the following comparison models:

- **Random:** This approach recommends n products randomly selected from all products. Measurements were performed 100 times and then averaged.
- **NLP-based Recommendation with Tokenization:** This model is the primary focus of our study, where product names are tokenized into individual tokens for analysis. The tokenization process involves breaking down complex product names into simpler, more manageable components, thereby allowing the system to learn and predict user preferences more accurately. The model utilizes the same hyperparameters as the non-tokenized version for a controlled comparison.
- **NLP-based Recommendation without Tokenization:** As a contrast to our primary model, this version of the recommendation system processes product names as whole entities, without breaking them down into tokens. This model serves as a direct comparison to evaluate the added value of tokenization in improving recommendation accuracy.
- **NLP-based Recommendation using N-grams:** In addition to the above models, we introduced a variant that utilizes n-gram tokenization. This model was tested with unigrams, bigrams, and trigrams to assess how different levels of token granularity impact the system's performance. Similar to other models, this approach also uses the same set of hyperparameters to ensure consistency in the comparative analysis.

The selection of these specific models was driven by the need to comprehensively evaluate the effectiveness of our tokenization approach against varying baselines. The Random model provides a baseline to contrast the predictive power of our NLP models against random chance. The NLP-based Recommendation without Tokenization model serves to directly assess the incremental benefit provided by tokenization. The inclusion of N-grams further allows us to explore the depth of tokenization granularity and its impact on recommendation accuracy. These comparative models were

chosen to provide a holistic understanding of our system's performance across different levels of text processing complexity.

In our experiment, we elaborate on the parameter settings utilized in our experiments to ensure transparency and reproducibility. The selection and tuning of these parameters were guided by preliminary tests and literature benchmarks to optimize the performance of our NLP-based recommendation system.

- **Transformer Model Configuration:** Our model utilized a Transformer architecture configured with a single layer ($\text{num_layers} = 1$), which was chosen to maintain a balance between complexity and computational efficiency. This layer count was found to be sufficient for capturing the nuances of our dataset while ensuring manageable training times.
- **Input and Output Space Dimensionality:** We set the dimensionality of the input and output space to 128 ($\text{d_model} = 128$). This dimension was selected as it provides a good trade-off between model expressiveness and overfitting risk, considering the size and complexity of our datasets.
- **Attention Mechanism:** The number of attention heads in the multi-head attention mechanism was set to 4 ($\text{num_heads} = 4$). This number allows the model to focus on different parts of the input sequence, improving its ability to learn from various patterns within the data.
- **Inner-Layer Dimensionality:** The inner-layer dimensionality was set to 256 (units = 256), which determines the capacity of the feed-forward networks within the Transformer. This setting was chosen to provide sufficient model complexity for learning intricate relationships in the data.
- **Dropout Rate:** To mitigate the risk of overfitting, we employed a dropout rate of 0.2 ($\text{dropout} = 0.2$) during training. This rate was optimized through cross-validation to ensure the model generalizes well to unseen data.
- **Training Epochs:** The model was trained across 100 epochs ($\text{epochs} = 100$), with early stopping applied to cease training upon convergence. This approach ensures that the model is adequately trained without overfitting to the training data.

These parameters were meticulously selected and adjusted to align with the specific needs and characteristics of our datasets, ensuring that our model delivers robust and reliable recommendations.

E. EXPERIMENTAL RESULTS

1) COMPARISON BETWEEN WITH AND WITHOUT TOKENIZATION

Our first focus was to emphasize the transformative impact of improving the overall efficiency of recommendation systems through token-level learning with product names. To ensure that our results were accurate and reliable, we utilized two

TABLE 3. Hit-Rate comparison for NLP-based recommendation without and with tokenization.

Datasets	Measure	Random	NLP-Based Recommendation		
			Without Tokenization	With Tokenization	Improvement (%)
UK e-Commerce	HR@3	0.0011	0.0340	0.0564	65.9%
	HR@5	0.0023	0.0487	0.0728	49.5%
	HR@10	0.0035	0.0690	0.0989	43.3%
	HR@15	0.0055	0.0953	0.1197	25.6%
	HR@20	0.0082	0.1115	0.1358	21.8%
Instacart	HR@3	0.0001	0.0205	0.0239	16.6%
	HR@5	0.0001	0.0311	0.0344	10.6%
	HR@10	0.0002	0.0461	0.0512	11.1%
	HR@15	0.0002	0.0532	0.0634	19.2%
	HR@20	0.0003	0.0588	0.0725	23.3%

TABLE 4. Mean Reciprocal Rank comparison for NLP-based recommendation without and with tokenization.

Datasets	Measure	Random	NLP-Based Recommendation		
			Without Tokenization	With Tokenization	Improvement (%)
UK e-Commerce	MRR@3	0.0011	0.0313	0.0407	30.0%
	MRR@5	0.0011	0.0360	0.0444	23.3%
	MRR@10	0.0012	0.0415	0.0478	15.2%
	MRR@15	0.0015	0.0438	0.0495	13.0%
	MRR@20	0.0015	0.0450	0.0504	12.0%
Instacart	MRR@3	0.0001	0.0138	0.0146	5.8%
	MRR@5	0.0001	0.0167	0.0170	1.8%
	MRR@10	0.0001	0.0190	0.0192	1.1%
	MRR@15	0.0001	0.0197	0.0202	2.5%
	MRR@20	0.0001	0.0201	0.0207	3.0%

key metrics: Hit-Rate and MRR. From Table 3, it is evident that infusing our recommendation system with a tokenization technique was not only beneficial but also provided a notable improvement. Taking the UK e-Commerce dataset as an example, when juxtaposed with its non-tokenized counterpart, the tokenized version exhibited a notable surge in performance. There is a significant enhancement rate, peaking at 65.9%. Similarly, experiments using the Instacart dataset yielded positive results. By employing tokenized product names in this particular dataset, we observed a surge in recommendation efficiency, with gains as high as 23.3%. To offer a holistic view of our results, we performed an in-depth examination of the Hit-Rate metric, placing special emphasis on the UK e-Commerce and Instacart datasets. Our insights from this are captured and visualized in Figure 5 for better clarity.

Moreover, MRR emerged as another pivotal metric in our study. This metric serves as a litmus test for recommendation systems and evaluating their performance based on the placement of the initial relevant recommendation. Specifically,

as portrayed in Table 4, it became clear that models trained on tokenized product names outperformed their peers in the UK e-Commerce dataset, showing an excellent improvement rate of up to 30.0%. Conversely, on the Instacart dataset, the performance enhancement of the tokenized model was more tempered, albeit still notable, showing an improvement of up to 5.8%. For a side-by-side comparison, both datasets and their respective performances are illustrated in Figure 6. Considering both the HR and MRR metrics, with a magnified focus on the UK e-Commerce and Instacart datasets, it was evident that while harnessing tokenized product names had an unmistakable edge in the UK e-Commerce landscape, it was less effective on the Instacart dataset. The latter presented a more reserved performance, possibly attributed to its inherently vast diversity in both user and product range, introducing a myriad of complexities into the learning curve of the recommendation system. Consolidating our research findings, it is evident that NLP-based recommendation systems, especially those enhanced by product name tokenization, provide compelling capability and efficiency. The strategic use of

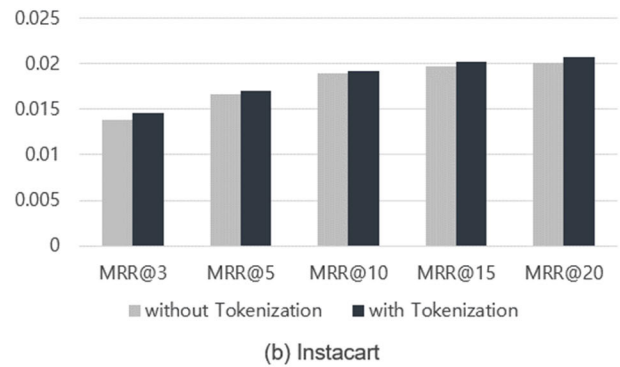
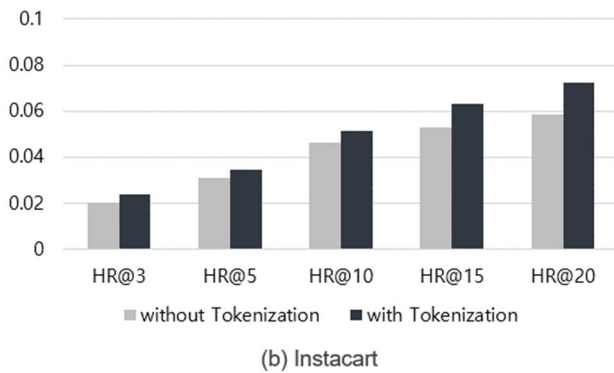
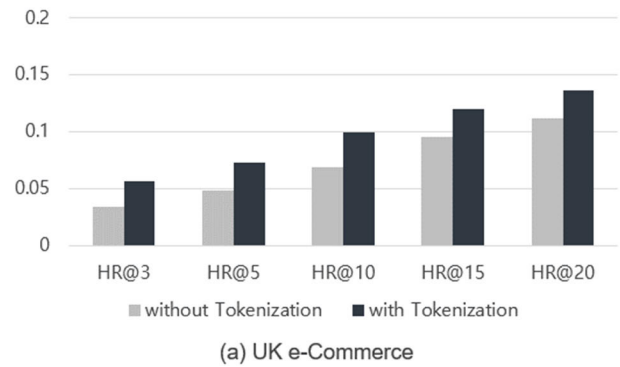
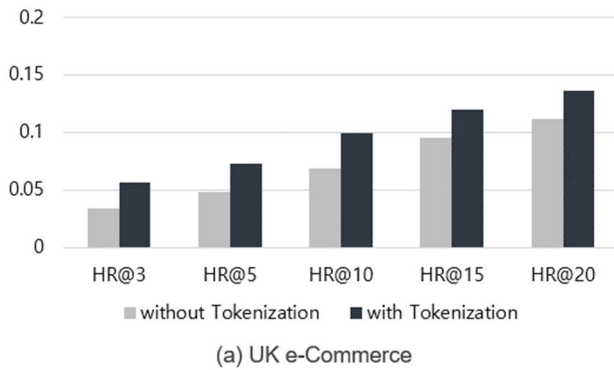


FIGURE 5. Hit-Rate comparison with and without tokenization.

tokenized product names not only provides benefits but also yields significant improvements in the system's performance. This favorable outcome remained consistent regardless of the dataset or evaluation metric applied.

2) COMPARISON OF TOKENIZATION ACROSS N-GRAMS

The second focal point of our experimentation was an in-depth analysis of the impact that various n-grams have on the performance of NLP-based recommendation systems. We scrutinized the efficiency of unigrams, bigrams, and trigrams in tokenizing product names to ascertain the most effective n-gram level for our tokenization strategy. Consistent with our methodology, we employed Hit-Rate and MRR metrics to measure accuracy. Initially, accuracy as depicted by the Hit-Rate in Table 5 demonstrated that both the UK e-Commerce and Instacart datasets achieved superior performance with unigram tokenization. The UK e-Commerce dataset showed a more noticeable variance in performance among the different n-grams, whereas the Instacart dataset exhibited less variation but nonetheless confirmed the superior performance of unigrams. This pattern is graphically represented in Figure 7, where unigrams take the lead, followed by bigrams and trigrams, in that order. The MRR metric mirrored these findings, with unigrams again showing the highest performance for both the UK e-Commerce and Instacart datasets, as corroborated by Table 6 and Figure 8. Across all our tests, unigrams consistently outperformed bigrams and

trigrams. This phenomenon is elucidated in Figure 9, which illustrates the breadth of interrelationships formed through unigram tokenization. Building on the premise that tokenization enables more detailed learning of product names, our n-grams experiment further solidified the finding that finer-grained tokenization facilitates more precise learning. Unigrams, by splitting product names into single-word units, created an extensive network of interrelationships that could be leveraged for a more comprehensive learning of patterns. Thus, the approach of tokenizing product names has been robustly validated through the n-grams experiment.

IV. SERVICE GENERATION

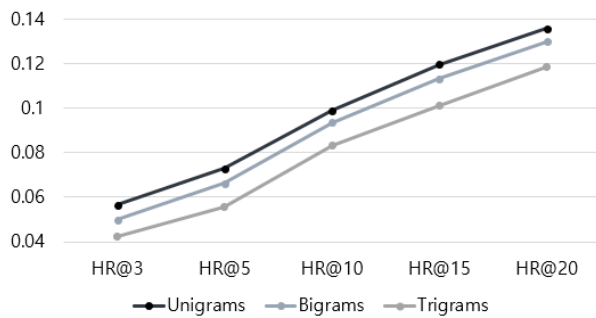
The previously introduced NLP-based recommendation system can derive product name tokens by learning them at the token level. As shown in Figure 10, this feature allows us to propose two services: New Product Brainstorming utilizing non-existent product names and Keyword Trend Forecasting using frequency analysis of product name tokens.

A. NEW PRODUCT BRAINSTORMING

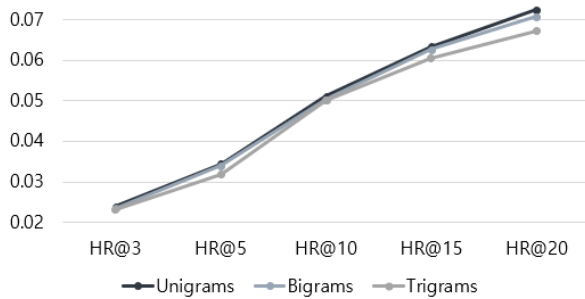
In initiating our discussion on the New Product Brainstorming service, we wish to underscore the exploratory and illustrative purpose of this application. The service is presented not as an endpoint, but as a demonstration of the NLP-based system's potential for broad service generation. It is a proof of concept that invites stakeholders to visualize the

TABLE 5. Hit-Rate comparison by n-grams.

Datasets	Measure	Product Name with Tokenization		
		Unigrams	Bigrams	Trigrams
UK e-Commerce	HR@3	0.0564	0.0499	0.0421
	HR@5	0.0728	0.0663	0.0557
	HR@10	0.0989	0.0937	0.0832
	HR@15	0.1197	0.1133	0.1010
	HR@20	0.1358	0.1299	0.1185
Instacart	HR@3	0.0239	0.0235	0.0232
	HR@5	0.0344	0.0340	0.0319
	HR@10	0.0512	0.0513	0.0505
	HR@15	0.0634	0.0627	0.0604
	HR@20	0.0725	0.0708	0.0674



(a) UK e-Commerce



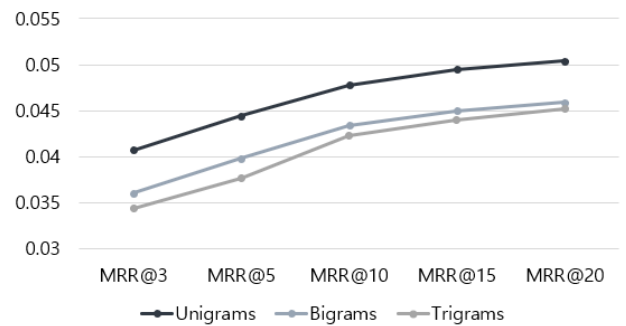
(b) Instacart

FIGURE 7. Hit-Rate comparison by n-grams.

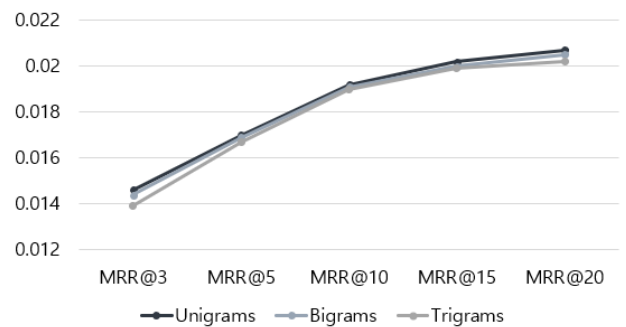
possibilities of NLP in product innovation. This aligns with our wider objective of highlighting the system's versatility and serves as a catalyst for further creative exploration and development. While product names that do not exist in reality might be seen as mere mistakes or errors, they can be utilized to generate novel product name ideas through "Serendipity," which refers to valuable discoveries or inventions made unintentionally through coincidences [43]. Particularly in the field of scientific research, there are many instances in which significant findings arise from experimental failure. A classic example is the discovery of penicillin by Fleming [44]. By mistakenly mixing in blue mold during a culture experiment, Fleming discovered an antimicrobial substance that

TABLE 6. MRR comparison by n-grams.

Datasets	Measure	Product Name with Tokenization		
		Unigrams	Bigrams	Trigrams
UK e-Commerce	MRR@3	0.0407	0.0360	0.0344
	MRR@5	0.0444	0.0398	0.0376
	MRR@10	0.0478	0.0434	0.0413
	MRR@15	0.0495	0.0450	0.0435
	MRR@20	0.0504	0.0459	0.0450
Instacart	MRR@3	0.0146	0.0129	0.0160
	MRR@5	0.0170	0.0150	0.0184
	MRR@10	0.0192	0.0179	0.0209
	MRR@15	0.0202	0.0190	0.0219
	MRR@20	0.0207	0.0196	0.0225



(a) UK e-Commerce



(b) Instacart

FIGURE 8. MRR comparison by n-grams.

was effective against infections. Other examples include the discovery of microwaves from melted chocolate and the birth of Post-it notes from a failed strong adhesive [45]. Based on these scientific cases, we aim to propose product names that do not exist in reality but are derived through the NLP-based recommendation system, as a foundation for brainstorming new product ideas.

1) UK E-COMMERCE DATA

From a test set of 7,121 rows, we generated the top-20 product names, resulting in a total of 142,420 product names. By comparing the product names in the dataset with the generated

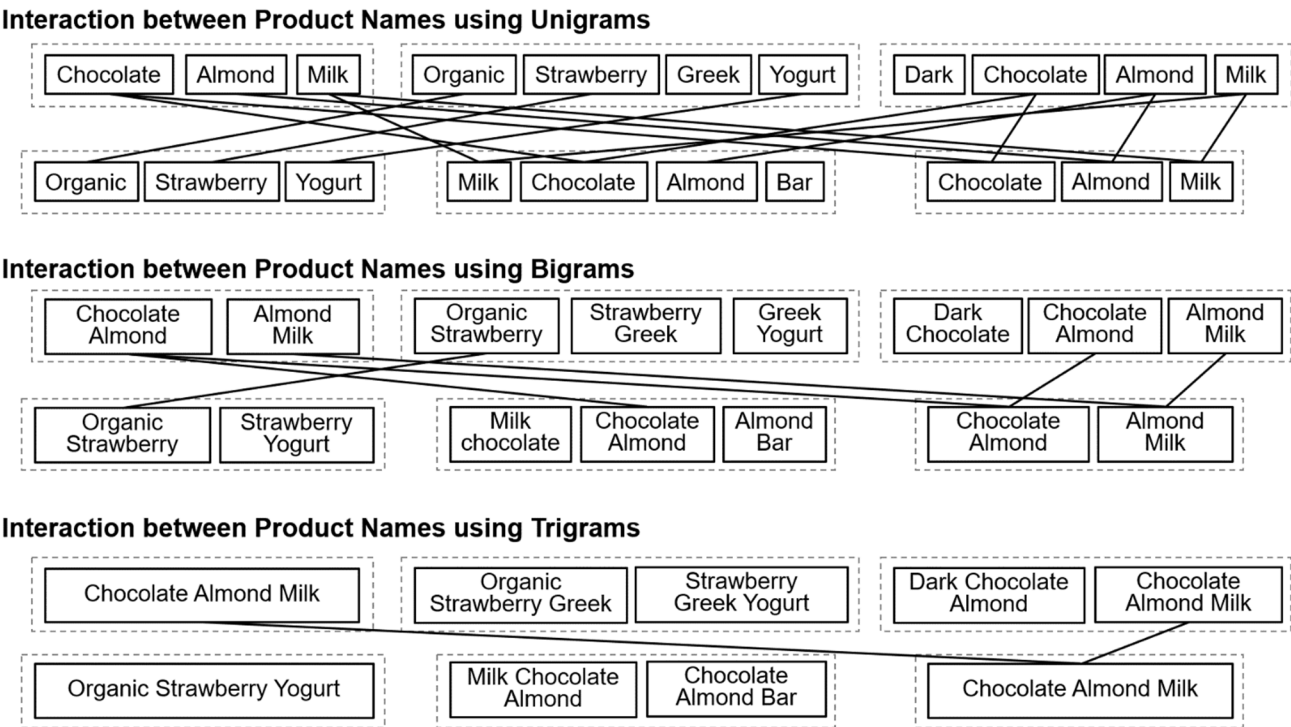


FIGURE 9. Interactions among product names by n-grams.

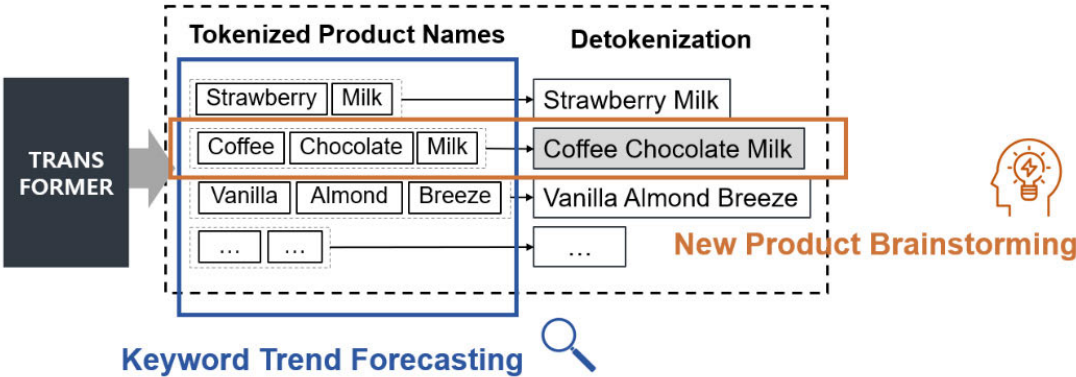


FIGURE 10. Service generation with NLP-based recommendation.

names, we found that 202 (0.15%) product names were not present in the dataset. Examples of non-existing product names are displayed on the left side of Figure 11. On the right-hand side, the most similar product names derived using Jaccard similarity are listed. Some product names had inaccurately generated tokens related to size or color, resulting in outputs like ‘Red Spotty Luggage Tag’ instead of ‘Pink Spotty Luggage Tag’ and ‘Green Owl Soft Toy’ instead of ‘Pink Owl Soft Toy’. Although ‘Red Spotty Luggage Tag’ and ‘Green Owl Soft Toy’ are not present in the dataset, they are colors that customers might desire. Additionally, there were cases where product name tokens were incorrectly generated or entirely new product names emerged, such as ‘Tube Red Spotty Paper Plates’, ‘Black Greeting Card Holder’, and ‘Tutti Frutti Notebook Box’. Even if similar products

are available in other stores, they can be considered novel product ideas based on the current store’s data. Specifically, to better understand unfamiliar names like ‘Tube Red Spotty Paper Plates’, we utilized DALL-E 3³ to generate images of the aforementioned product names, as depicted in Figure 12. In the future, by expanding the “new product brainstorming” service, we aim to offer new product development ideas by presenting both product names and images together.

2) INSTACART DATASET

From a test set of 16,644 rows, we generated the top-20 product names, resulting in a total of 332,880 product names. By comparing the product names in the dataset with the

³<https://openai.com/dall-e-3>

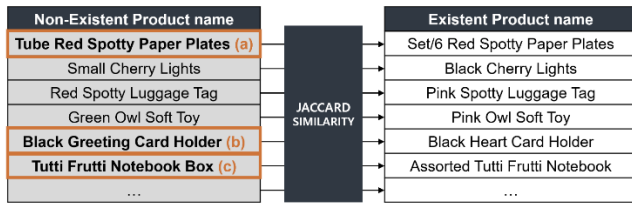


FIGURE 11. New product brainstorming based on the UK e-Commerce dataset.

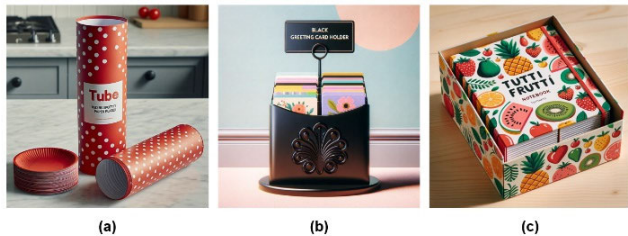


FIGURE 12. Expected images of new product brainstorming based on the UK e-Commerce dataset: (a) Tube Red Spotty Paper Plates (b) Black Greeting Card Holder (c) Tutti Frutti Notebook Box.

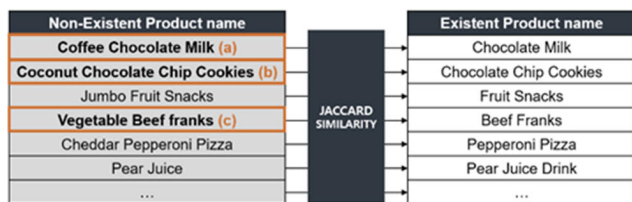


FIGURE 13. New product brainstorming based on Instacart dataset.



FIGURE 14. Expected images of new product brainstorming based on the Instacart dataset: (a) Coffee Chocolate Milk (b) Coconut Chocolate Chip Cookies (c) Vegetable Beef franks.

generated names, we found that 1,020 (0.31%) product names were not present in the dataset. Examples of non-existing product names are displayed on the left side of Figure 13. On the right-hand side, the most similar product names, derived using Jaccard similarity, are listed. Some product names had inaccurately generated tokens related to flavor or ingredients, resulting in outputs like ‘Coffee Chocolate Milk’ instead of ‘Chocolate Milk’ and ‘Coconut Chocolate Chip Cookies’ instead of ‘Chocolate Chip Cookies’. Although ‘Coffee Chocolate Milk’ and ‘Coconut Chocolate Chip Cookies’ are not present in the dataset, they might represent flavors or variations that are appealing to customers. There were also

instances in which product name tokens were generated in a different manner or entirely unique product names surfaced, such as ‘Vegetable Beef Franks’. Even if such products exist in other markets or stores, they can be viewed as fresh product concepts based on the current store’s data. For a clearer visualization of these potential products, we used DALL·E 3 to generate images corresponding to the product names, as shown in Figure 14. Product names that do not exist in reality and are derived from a token can be used for new product brainstorming, serving as a support service for the product development department and company.

B. KEYWORD TREND FORECASTING

Several studies have analyzed trends through keyword analysis, either by analyzing the frequency of words in a text to identify research trends or by predicting product sales and consumer purchase trends using product frequency analysis [46], [47]. The results of a recommendation system can be viewed as data that predict what users will purchase in the future based on their past purchases. In particular, the proposed NLP-based recommendation method outputs product names in token units; thus, these tokens can be considered as keywords of product elements. Therefore, we suggest a keyword trend forecasting service based on a frequency analysis of product keywords.

1) UK E-COMMERCE DATA

Based on UK e-commerce data, an experiment was conducted using 7,121 test data entries to generate the top 20 product names. Consequently, a total of 142,420 product names were created. These product names were tokenized, and a frequency analysis was conducted based on these tokens. The results are presented in Table 7, where top keywords such as ‘Heart’, ‘Red’, and ‘Set’ were identified. Each entry in this test set has four data points representing a user’s past purchases and one label indicating the next expected purchase. The four learning data points represent a user’s purchase history, while the single label indicates the product that the user is expected to purchase next. By analyzing the frequency of all products corresponding to this label, we can predict which products will be popular in the future. The results are shown in Table 8, where notably, the ‘White Hanging Heart T-light Holder’ product appeared with the highest frequency. Combining the results in Tables 7 and 8, we observed that products containing specific keywords have high sales volumes. For instance, products with the keywords ‘Heart’, ‘White’, and ‘Hanging’ show significant sales. This suggests that keyword-based trend forecasting can significantly influence product sales strategies. In Figure 15, the top-50 keywords are visualized as a word cloud.

2) INSTACART DATA

Based on e-commerce data from Instacart, a test was conducted using a total of 16,644 rows of data to generate the top 20 product names. This resulted in 332,880 product

TABLE 7. Top-15 keywords from NLP-based recommendation on the UK e-Commerce dataset.

RANK	Keyword	#Keywords	#Products with Keyword	Product%
1	Heart	16,919	99	1.4%
2	Red	14,341	126	1.8%
3	Set	13,380	127	1.8%
4	Bag	11,936	88	1.2%
5	Spotty	11,455	61	0.9%
6	Pink	9,618	149	2.1%
7	Hanging	9,340	57	0.8%
8	Holder	9,269	55	0.8%
9	White	8,696	74	1.0%
10	Cake	7,986	62	0.9%
11	Metal	7,039	65	0.9%
12	Blue	6,844	111	1.6%
13	T-light	6,736	42	0.6%
14	Small	6,309	62	0.9%
15	Box	6,210	87	1.2%

TABLE 8. Top-15 products from Label in the UK e-Commerce dataset.

RANK	Product Name	#	%
1	White Hanging Heart T-light Holder	44	0.6%
2	Pack of 72 Retro Spot Cake Cases	34	0.5%
3	Door Mat Union Flag	34	0.5%
4	Love Building Block Word	31	0.4%
5	Home Building Block Word	30	0.4%
6	Regency Cakestand 3 Tier	26	0.4%
7	Set/20 Red Spotty Paper Napkins	24	0.3%
8	Strawberry Ceramic Trinket Box	22	0.3%
9	60 Teatime Fairy Cake Cases	22	0.3%
10	Lunch Bag Red Spotty	21	0.3%
11	Jumbo Bag Pink with White Spots	21	0.3%
12	Set/5 Red Spotty Lid Glass Bowls	21	0.3%
13	Round Snack Boxes Set of 4 Woodland	20	0.3%
14	Wooden Frame Antique White	20	0.3%
15	Paper Chain Kit Retro Spot	20	0.3%

names being generated. The produced product names were tokenized, and a frequency analysis was conducted based on these tokens. The results are presented in Table 9, where the top product names, such as ‘Banana’, ‘Bag of Organic Bananas’, and ‘Organic Strawberries’, were identified. Each row of this test set consists of four learning data points and one label. The four learning data points represent a user’s purchase history, while the single label indicates the product the user is expected to purchase next. Analyzing the frequency of all products corresponding to this label allows us to predict which products may be popular in the future. This result is shown in Table 10, where notably, the keyword ‘organic’ appeared with the highest frequency. Combining the results in Tables 9 and 10, we observed that



FIGURE 15. Word cloud of the top-50 keywords from NLP-based recommendation on the UK e-Commerce dataset.

TABLE 9. Top-15 keywords from NLP-based recommendation on the Instacart dataset.

RANK	Keyword	#Keywords	#Products with Keyword	Product%
1	organic	37,937	5,253	31.6%
2	milk	26,280	848	5.1%
3	whole	25,436	635	3.8%
4	banana	23,204	501	3.0%
5	yogurt	21,855	648	3.9%
6	original	18,048	525	3.2%
7	almond	17,510	389	2.3%
8	hummus	16,277	112	0.7%
9	water	16,095	524	3.1%
10	red	15,781	369	2.2%
11	free	15,108	581	3.5%
12	bread	14,931	292	1.8%
13	chicken	14,927	417	2.5%
14	pepper	14,588	250	1.5%
15	natural	13,485	326	2.0%

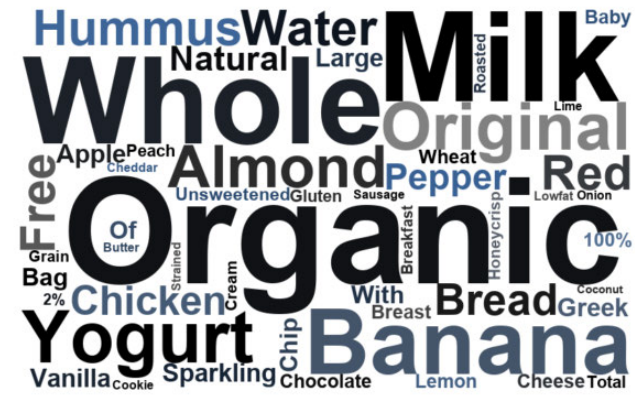


FIGURE 16. Word cloud of the top-50 keywords from NLP-based recommendation on the Instacart dataset.

products containing specific keywords have high sales. For instance, products containing the keywords ‘organic’, ‘milk’,

TABLE 10. Top-15 products from label in the Instacart dataset.

RANK	Product Name	#	%
1	Banana	235	1.4%
2	Bag of Organic Bananas	190	1.1%
3	Organic Strawberries	155	0.9%
4	Organic Baby Spinach	116	0.7%
5	Organic Hass Avocado	116	0.7%
6	Organic Avocado	100	0.6%
7	Strawberries	85	0.5%
8	Organic Whole Milk	81	0.5%
9	Limes	74	0.4%
10	Large Lemon	66	0.4%
11	Organic Garlic	62	0.4%
12	Organic Raspberries	62	0.4%
13	Organic Blueberries	54	0.3%
14	Organic Zucchini	52	0.3%
15	Organic Yellow Onion	50	0.3%

and ‘whole’ showed significant sales volumes. This suggests that keyword-based trend forecasting can be instrumental in product sales strategies. In Figure 16, the top-50 keywords are visualized as a word cloud.

V. CONCLUSION

The main finding of this study is the demonstrated efficacy of token-level analysis in NLP-based recommendation systems, leading to significant performance improvements in the context of e-commerce. Additionally, our exploration into n-grams, particularly the superior performance of unigram tokenization, further reinforces the effectiveness of fine-grained token analysis in enhancing recommendation system accuracy. The n-grams experiment, especially the comparative analysis of unigrams, bigrams, and trigrams, provided crucial insights into how different granularities of tokenization affect the predictive accuracy and utility of our NLP-based recommendation system, thus contributing to a deeper understanding of the nuances in natural language processing for e-commerce applications. In the modern e-commerce landscape, understanding consumer behavior requires the analysis of vast amounts of data, with particular emphasis on natural language data. In this study, we adopted a unique approach by utilizing product names at the token level, thereby experimentally validating the efficacy of this methodology. Using the UK e-Commerce and Instacart datasets, we measured the performance of an NLP-based recommendation algorithm that employs product names at the token level. The results confirmed its high efficacy. This underscores the potential of NLP technologies to move beyond mere data analysis and emphasize deep data insights and the provision of high-quality services. Furthermore, this study highlighted the significance of creating NLP-based services with a focus on offering services without the use of personal data. Our natural language learning approach, which

analyzes product names at the token level, has demonstrated its potential to offer effective services while preserving user privacy. By focusing solely on product name tokens, rather than personal user data, we enhance the system’s privacy and mitigate concerns related to personal data use. The innovative services we introduced, specifically product brainstorming and keyword trend forecasting, stem directly from our findings that token-level analysis of product names can uncover latent consumer preferences and market trends. These services hold the potential to revolutionize business strategies by providing insights into untapped product opportunities and emerging market demands. Future research should explore the scalability of token-level analysis across larger and more diverse datasets, and examine the integration of evolving NLP technologies to maintain the relevance and effectiveness of recommendation systems.

From a theoretical perspective, this study makes two significant contributions. First, it substantially improves the performance of recommendation algorithms. By advancing existing NLP-based recommendation methods and incorporating tokenization of product names, our approach demonstrated superior performance, as evidenced by improved metrics such as Hit-Rate and Mean Reciprocal Rank in comparison to non-tokenized models. Second, notable progress was made in the extensibility of the recommendation model. NLP-based recommendation models can be flexibly applied across diverse languages and domains, thereby enabling services to be offered globally, unrestricted by specific languages or regions. This study had two important practical implications. First, processing product names in natural language eliminates the need for personal information, thereby significantly enhancing their potential for industrial applications. This method can be integral in industries such as e-commerce and digital marketing, where it can enhance user experience by providing personalized recommendations without compromising individual privacy, thereby addressing challenges related to the protection of personal information. Second, this study explored the potential of diverse services. The development of services, such as brainstorming new product ideas and predicting keyword trends, offers businesses a sustainable path for utilization. Thus, companies are better equipped to respond swiftly to market shifts and evolving consumer demand.

Despite offering significant insights, this study has certain limitations that merit consideration. First, it was confined to the UK e-Commerce and Instacart datasets. Future research should consider applying our token-level analysis approach to a wider range of datasets, such as social media trends, customer reviews, and global marketplaces, to validate the applicability and reliability of our findings across various consumer contexts and cultural backgrounds. Second, the NLP technologies employed reflect the current state-of-the-art. As technology evolves, newer techniques may emerge, potentially affecting the performance and outcomes of recommendation algorithms. Furthermore, a notable advantage of this study lies in its approach to using product names directly

and analyzing them at the token level. However, this methodology is predominantly suitable for products with intuitive and straightforward names. For more abstract product names, the application could become challenging. For instance, for a product name like “Taste of Magic,” it is not immediately clear which food it represents. Thus, NLP methodologies may struggle to provide accurate recommendations based solely on such abstract names. Future research should focus on expanding the datasets tested, incorporating newer NLP techniques, and devising strategies to effectively handle products with abstract or non-descriptive names.

REFERENCES

- [1] E. Cambria and B. White, “Jumping NLP curves: A review of natural language processing research [review article],” *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 48–57, May 2014.
- [2] I. Lauriola, A. Lavelli, and F. Aioli, “An introduction to deep learning in natural language processing: Models, techniques, and tools,” *Neurocomputing*, vol. 470, pp. 443–456, Jan. 2022.
- [3] M. O. Topal, A. Bas, and I. van Heerden, “Exploring transformers in natural language generation: GPT, BERT, and XLNet,” 2021, *arXiv:2102.08036*.
- [4] A. Gillioz, J. Casas, E. Mugellini, and O. A. Khaled, “Overview of the transformer-based models for NLP tasks,” in *Proc. 15th Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2020, pp. 179–183.
- [5] O. S. Shalom, H. Roitman, and P. Kouki, “Natural language processing for recommender systems,” in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds. New York, NY, USA: Springer, 2021, pp. 447–483.
- [6] S. Lee, H. Shin, I. Choi, and J. Kim, “Analyzing the impact of components of Yelp.Com on recommender system performance: Case of Austin,” *IEEE Access*, vol. 10, pp. 128066–128076, 2022.
- [7] G. de Souza Pereira Moreira, S. Rabhi, J. M. Lee, R. Ak, and E. Oldridge, “Transformers4rec: Bridging the gap between NLP and sequential/session-based recommendation,” in *Proc. 15th ACM Conf. Recommender Syst.*, Sep. 2021, pp. 143–153.
- [8] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, “BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer,” in *Proc. 28th ACM Int. Conf. Inform. Knowl. Manage.*, Nov. 2019, pp. 1441–1450.
- [9] Q. Chen, H. Zhao, W. Li, P. Huang, and W. Ou, “Behavior sequence transformer for e-commerce recommendation in Alibaba,” in *Proc. 1st Int. Workshop Deep Learn. Pract. High-Dimensional Sparse Data*, Aug. 2019, pp. 1–4.
- [10] T. Perry, “What is Tokenization in natural language processing (NLP),” *Mach. Learn. Plus*, Jun. 2022. [Online]. Available: <https://www.machinelearningplus.com/nlp/what-is-tokenization-in-natural-language-processing>
- [11] K. Park, J. Lee, S. Jang, and D. Jung, “An empirical study of tokenization strategies for various Korean NLP tasks,” 2020, *arXiv:2010.02534*.
- [12] I. E. Tiffani, “Optimization of Naïve Bayes classifier by implemented Unigram, Bigram, Trigram for sentiment analysis of hotel review,” *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 1–7, 2020.
- [13] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, “From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy,” *IEEE Access*, vol. 11, pp. 80218–80245, 2023, doi: [10.1109/ACCESS.2023.3300381](https://doi.org/10.1109/ACCESS.2023.3300381).
- [14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, “Attention is all you need,” in *Proc. Adv. Neural Inform. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [15] H. Wang, J. Li, H. Wu, E. Hovy, and Y. Sun, “Pre-trained language models and their applications,” *Engineering*, vol. 25, pp. 51–65, Jun. 2022.
- [16] F. Zhang, G. An, and Q. Ruan, “Transformer-based natural language understanding and generation,” in *Proc. 16th IEEE Int. Conf. Signal Process. (ICSP)*, vol. 1, Oct. 2022, pp. 281–284.
- [17] B. Wang, Q. Huang, B. Deb, A. Halfaker, L. Shao, D. McDuff, and J. Gao, “Logical transformers: Infusing logical structures into pre-trained language models,” in *Proc. Findings Assoc. Comput. Linguist. (ACL)*, Jul. 2023, pp. 1762–1773.
- [18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” 2018, *arXiv:1810.04805*.
- [19] G. Yenduri, G. Srivastava, P. K. R. Maddikunta, R. H. Jhaveri, W. Wang, A. V. Vasilakos, and T. R. Gadekallu, “Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions,” 2023, *arXiv:2305.10435*.
- [20] X. Yu, F. Jiang, J. Du, and D. Gong, “A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains,” *Pattern Recognit.*, vol. 94, pp. 96–109, Oct. 2019.
- [21] X. Yu, Q. Peng, L. Xu, F. Jiang, J. Du, and D. Gong, “A selective ensemble learning based two-sided cross-domain collaborative filtering algorithm,” *Inf. Process. Manage.*, vol. 58, no. 6, Nov. 2021, Art. no. 102691.
- [22] F. Huang, Z. Wang, X. Huang, Y. Qian, Z. Li, and H. Chen, “Aligning distillation for cold-start item recommendation,” in *Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2023, pp. 1147–1157.
- [23] Y. Lu, K. Nakamura, and R. Ichise, “HyperRS: hypernetwork-based recommender system for the user cold-start problem,” *IEEE Access*, vol. 11, pp. 5453–5463, 2023.
- [24] C. Channarong, C. Paosirikul, S. Maneeroj, and A. Takasu, “Hybrid-BERT4Rec: A hybrid (content-based filtering and collaborative filtering) recommender system based on BERT,” *IEEE Access*, vol. 10, pp. 56193–56206, 2022.
- [25] X. Liu, G. Zhou, M. Kong, Z. Yin, X. Li, L. Yin, and W. Zheng, “Developing multi-labelled corpus of Twitter short texts: A semi-automatic method,” *Systems*, vol. 11, no. 8, p. 390, Aug. 2023.
- [26] X. Liu, S. Wang, S. Lu, Z. Yin, X. Li, L. Yin, J. Tian, and W. Zheng, “Adapting feature selection algorithms for the classification of Chinese texts,” *Systems*, vol. 11, no. 9, p. 483, Sep. 2023.
- [27] K. J. Lee, Y. Hwangbo, H. Jung, B. Jeong, and J. I. Park, “TransformRec: User-centric recommender system for e-commerce using transformer,” in *Proc. 23rd Int. Center Electron. Commerce*, 2022, pp. 171–178.
- [28] D. Sangiorgi and Y. Eun, “Service design as an approach to new service development: Reflections and future studies,” in *Proc. ServDes. Service Future, Proc. 4th Service Design Service Innov. Conf., Linköping Electron. Conf.*, 2014, pp. 194–204.
- [29] J. Wang, H. Guo, and J. Chen, “Research on extension innovation model in the creation process of service design,” *Proc. Comput. Sci.*, vol. 199, pp. 992–999, 2022.
- [30] S. Pais, J. Cordeiro, and M. L. Jamil, “NLP-based platform as a service: A brief review,” *J. Big Data*, vol. 9, no. 54, pp. 1–26, 2022, doi: [10.1186/s40537-022-00603-5](https://doi.org/10.1186/s40537-022-00603-5).
- [31] A. Copestake, *Natural Language Processing: Part 1 of Lecture Notes (Ann Copestake Lecture Note Series)*, Cambridge, U.K., 2003.
- [32] D. M. Zajic, B. J. Dorr, and J. Lin, “Single-document and multi-document summarization techniques for email threads using sentence compression,” *Inf. Process. Manage.*, vol. 44, no. 4, pp. 1600–1610, Jul. 2008.
- [33] M. A. Fattah and F. Ren, “GA, MR, FFNN, PNN and GMM based models for automatic text summarization,” *Comput. Speech Lang.*, vol. 23, no. 1, pp. 126–144, Jan. 2009.
- [34] L. Fan, L. Li, Z. Ma, S. Lee, H. Yu, and L. Hemphill, “A bibliometric review of large language models research from 2017 to 2023,” 2023, *arXiv:2304.02020*.
- [35] M. Liao and S. S. Sundar, “When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering,” *J. Advertising*, vol. 51, no. 2, pp. 256–267, Mar. 2022.
- [36] K. J. Lee, B. Jeong, Y. Hwangbo, S. Kim, and M. Yang, “Target marketing methodology as a dual of user-centric hyper-personalized recommendation systems for multi-merchant environments,” in *Proc. Korea Soc. Manag. Inf. Syst. Joint Spring Conf.*, Busan, South Korea, Jun. 2022.
- [37] Q. Deng, K. Wang, M. Zhao, Z. Zou, R. Wu, J. Tao, and L. Chen, “Personalized bundle recommendation in online games,” in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2381–2388.
- [38] M. Beladev, L. Rokach, and B. Shapira, “Recommender systems for product bundling,” *Knowl.-Based Syst.*, vol. 111, pp. 193–206, Nov. 2016.
- [39] S. Kim, B. Jeong, G. Lee, Y. Cho, and K. J. Lee, “Product bundling and co-marketing methodology through big data analysis of recommendation system results,” in *Proc. Korea Intell. Inf. Syst. Soc. Fall Conf.*, Seoul, South Korea, Nov. 2022.
- [40] M. Chen and P. Liu, “Performance evaluation of recommender systems,” *Int. J. Performability Eng.*, vol. 13, no. 8, p. 1246, 2017.

- [41] I. Palomares, C. Porcel, L. Pizzato, I. Guy, and E. Herrera-Viedma, "Reciprocal recommender systems: Analysis of state-of-art literature, challenges and opportunities towards social recommendation," *Inf. Fusion*, vol. 69, pp. 103–127, May 2021.
- [42] A. Gunawardana, G. Shani, and S. Yogev, "Evaluating recommender systems," in *Recommender Systems Handbook*. New York, NY, USA: Springer, 2012, pp. 547–601.
- [43] J. Corneli, A. Jordanous, C. Guckelsberger, A. Pease, and S. Colton, "Modelling serendipity in a computational context," 2014, *arXiv:1411.0440*.
- [44] J. W. Bennett and K. T. Chung, "Alexander fleming and the discovery of penicillin," in *Advances in Applied Microbiology*, vol. 49. Cambridge, MA, USA: Academic, 2001.
- [45] P. Andriani, "Exaptation, serendipity and aging," *Mech. Ageing Develop.*, vol. 163, pp. 30–35, Apr. 2017.
- [46] N. A. Mazov, V. N. Gureev, and V. N. Glinskikh, "The methodological basis of defining research trends and fronts," *Sci. Tech. Inf. Process.*, vol. 47, no. 4, pp. 221–231, Oct. 2020.
- [47] F. Dotsika and A. Watkins, "Identifying potentially disruptive trends by means of keyword network analysis," *Technol. Forecasting Social Change*, vol. 119, pp. 114–127, Jun. 2017.



BAEK JEONG received the B.S. and M.S. degrees in child and family studies from Kyung Hee University, Seoul, South Korea, where he is currently pursuing the Ph.D. degree in big data analytics. Since 2020, he has been a Senior Researcher with the Artificial Intelligence and Business Models (AI & BM) Laboratory. His research interests include natural language processing, generative AI, and youth development. He was a recipient of the Innovative Applications of Artificial Intelligence (IAAI) Award from the Association for the Advancement of Artificial Intelligence (AAAI), in 2024.



KYOUNG JUN LEE received the B.S., M.S., and Ph.D. degrees in management science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, and the degree in public administration from Seoul National University, Seoul, South Korea. He was a Visiting Scientist/Professor at CMU, MIT, and UC Berkeley. In 2017, he was the President of the Korean Intelligent Information Systems Society. He founded Benple Inc., an IoT & AI-based company, and Allwin Inc., a patent-based e-commerce company. He is currently a Professor of business administration and big data analytics and the Director of the Artificial Intelligence and Business Models (AI & BM) Laboratory and the Big Data Research Center, Kyung Hee University, Seoul. His contribution to Korea's e-government system earned him the Presidential Award, in 2018. He was a recipient of the Innovative Applications of Artificial Intelligence (IAAI) Awards from the Association for the Advancement of Artificial Intelligence (AAAI), in 1995, 1997, 2020, and 2024.

...