Impact Factor (SJIF): 6.126 E-ISSN: 2583-1615, P-ISSN: 3049-3498

International Journal of Research in Academic World

Received: 08/April/2025 IJRAW: 2025; 4(5):188-192 Accepted: 11/May/2025

Product Recommendation and Product Matching Using Deep Learning

*1Dr. Gandhimathi K, 2Savitha B and 3Shrinithi V

*1Assistant Professor, Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu, India.

2, 3PG Student, Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu.

Abstract

The swift expansion of e-commerce has rendered product selection increasingly intricate, influenced by factors such as pricing, customer feedback, and ratings. This project introduces a recommendation system based on deep learning, aimed at aiding users in identifying the most appropriate product within a specified category. By utilizing structured data, which includes price, ratings, and sales channels, alongside unstructured customer reviews, the system enhances the precision of its recommendations. The dataset is subjected to preprocessing, where categorical variables are transformed through label encoding, and numerical features are standardized using StandardScaler. Customer reviews are processed through tokenization and converted into padded sequences for subsequent text analysis. The model effectively combines both numerical and textual data, ensuring a holistic representation of products. A deep learning framework is constructed using TensorFlow and Keras, featuring multiple dense layers augmented with batch normalization and dropout techniques to enhance generalization and mitigate overfitting. Training is conducted using sparse categorical cross-entropy loss, with optimization performed via the Adam algorithm. The efficacy of the system is assessed through metrics such as accuracy, R^2 scores. The model achieves competitive accuracy, with further validation through loss minimization to gauge overall performance. This initiative underscores the capabilities of deep learning in the realm of e-commerce recommendations, facilitating users in making informed purchasing choices based on data. By incorporating both numerical and textual attributes, it enriches the personalized shopping experience, thereby fostering a more effective and competitive e-commerce environment. The suggested framework illustrates how deep learning techniques can optimize product selection, ultimately serving the interests of both consumers and online retailers by providing accurate and dependable recommendations within an expanding digital marketplace.

Keywords: Product, E-commerce, Neural network, R2 Score, Recommendation, Product Matching.

1. Introduction

E-commerce offers consumers a vast selection of products, which can complicate the process of identifying the most suitable option based on factors such as price, reviews, and ratings. This project seeks to create a recommendation system powered by deep learning that combines both structured data (including pricing, ratings, and sales channels) and unstructured data (such as customer reviews) to facilitate better decision-making. The implementation of advanced methodologies, including feature encoding, normalization, and natural language processing techniques like sentiment analysis and named entity recognition (NER), is essential for deriving valuable insights. The model utilizes neural networks to process the data and deliver precise recommendations. Its effectiveness is assessed through various metrics, including accuracy and the R2 score. This system simplifies the product selection process, thereby conserving consumers' time and enhancing their satisfaction through tailored suggestions. By integrating AI-driven recommendations, e-commerce platforms can boost user engagement and sales, establishing a new benchmark for intelligent shopping assistance in a rapidly changing digital landscape.

2. Literature Review

Li *et al*. (2023) review deep learning-based recommendation systems, highlighting their progress, applications, and challenges. The study categorizes recommendation strategies into collaborative filtering, content-based filtering, and hybrid approaches, emphasizing the role of CNNs, RNNs, LSTMs, and transformers in improving recommendation accuracy. It explores NLP and attention mechanisms for better user intent understanding. Key challenges include data sparsity, computational demands, and cold start issues, with suggestions for future advancements. This review provides valuable insights into deep learning's impact on personalized recommendations, benefiting both academia and industries like e-commerce and entertainment [1].

Shanthi *et al*. (2023) explore a deep learning-based product recommendation system using Recurrent Neural Networks (RNNs) to enhance personalized recommendations. The study highlights the limitations of traditional methods like collaborative and content-based filtering, which struggle with cold start issues and data sparsity. By leveraging deep learning, the system analyzes sequential user interactions, textual reviews, and purchase history. Embedding layers and

*Corresponding Author: Dr. Gandhimathi K

attention mechanisms further improve accuracy and relevance. The findings demonstrate significant improvements in prediction precision and user satisfaction, with future research suggesting hybrid models and reinforcement learning to refine recommendation quality in e-commerce and online retail [2].

Zhang *et al*. (2025) explore a deep learning-based personalized product recommendation model for AI-driven question-and-answer robots. The study focuses on NLP and recommendation algorithms to enhance user engagement by analyzing inquiries, preferences, and past interactions. Using deep neural networks (DNNs), recurrent neural networks (RNNs), and attention mechanisms, the model captures contextual relationships for precise recommendations. It addresses challenges like data sparsity and real-time processing through hybrid systems and reinforcement learning. Experimental results show improved recommendation accuracy and user satisfaction, highlighting the model's potential for e-commerce, virtual assistants, and intelligent customer service platforms [3].

Tayade *et al*. (2021) explore deep learning in product recommendation systems, emphasizing its role in enhancing user experience and decision-making. The study examines CNNs, RNNs, and autoencoders, comparing them to traditional methods like collaborative and content-based filtering, which struggle with cold start issues and data sparsity. By leveraging user behavior analysis, sentiment extraction, and feature embeddings, the approach delivers personalized recommendations. Applications span e-commerce, streaming platforms, and digital marketing. Results show that hybrid models combining deep learning with traditional techniques improve accuracy, paving the way for more intelligent and adaptive recommendation systems [4].

Xu, He *et al*. (2018) examine the impact of deep learning on search and recommendation matching mechanisms, highlighting its role in improving relevance and user satisfaction. Traditional techniques like keyword searches and matrix factorization struggle with semantic understanding and complex preferences. By leveraging CNNs, RNNs, and attention mechanisms, the proposed models capture high-dimensional feature representations for more precise matching. The study integrates representation learning and interaction-based models to enhance context comprehension. Findings show that deep learning significantly improves ranking performance, enabling the development of more intelligent and adaptive information retrieval systems for search and recommendation applications [5].

Yulianton *et al*. explore product matching in e-commerce, addressing the challenge of identifying identical products with varying titles and descriptions. Traditional methods struggle with data complexity, but transformer-based models like BERT have improved performance. Tracz *et al*. (2020) demonstrated BERT's effectiveness in similarity learning, while Yulianton and Santi (2024) advanced this with Sentence-BERT (SBERT) for product deduplication. Using the all-MiniLM-L6-v2 variant, their approach integrates text preprocessing, strategic training pairs, and threshold-based similarity matching. Evaluated on the Pricerunner dataset, their model achieved 98.10% accuracy, highlighting SBERT's efficiency in enhancing e-commerce product matching and deduplication [6].

3. Domain–Ecommerce

E-commerce, or electronic commerce, refers to the buying and selling of goods, services, or digital products via the internet. This innovative approach has transformed the conventional retail landscape by allowing businesses to connect with a worldwide customer base, operate continuously, and deliver a streamlined shopping experience. E-commerce encompasses various models, including Business-to-Consumer (B2C), Business-to-Business (B2B), Consumer-to-Consumer (C2C), and Consumer-to-Business (C2B), each designed to facilitate distinct transaction types. E-commerce has fundamentally altered the dynamics between businesses and consumers, providing unmatched convenience, diversity, and accessibility. As technological advancements and digital infrastructure continue to evolve, the sector is poised for further growth, incorporating innovations such as AI, blockchain, and AR/VR. Companies must remain agile in response to shifting consumer behaviors and technological development.

4. Methodology

i). Data Collection

The dataset was collected from multiple e-commerce platforms, including Amazon and Flipkart, using web scraping techniques. Python libraries such as requests and BeautifulSoup were utilized for data extraction, while APIs were leveraged where available. The collected dataset included structured data (e.g., price, ratings, product categories) and unstructured data (e.g., customer reviews). Data was stored in a structured format such as CSV for further processing. To ensure data accuracy and completeness, duplicate records were identified and removed. Data consistency was maintained by standardizing the format of product names, pricing, and category labels. The collected data was stored in a relational database to facilitate efficient retrieval and analysis.

ii). Workflow

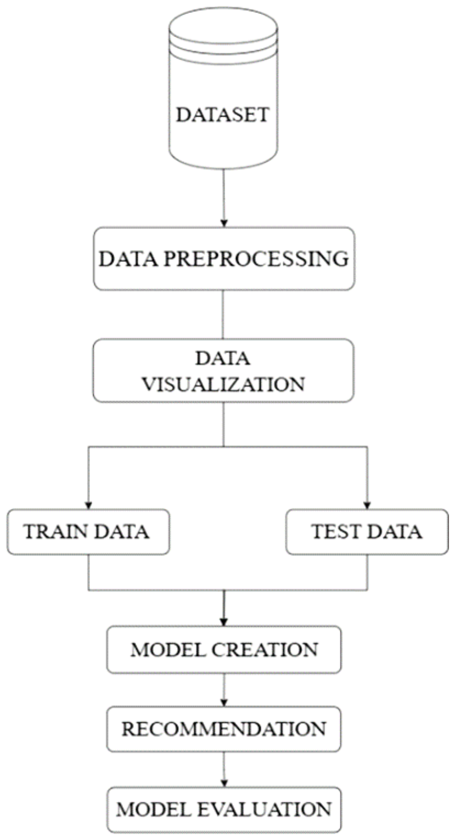

Fig 1: Workflow of the Proposed Work

iii). Dataset

The dataset consists of product listings from multiple e-commerce platforms, including Amazon, Flipkart, Meesho, Myntra, Nykaa, and Ajio. It includes various product categories such as home appliances, beauty products, clothing, footwear, and electronics. Each entry in the dataset contains a unique Product ID, a Title, and a Description summarizing the product's key features. The datas*et al*so includes Price information, expressed in Indian Rupees (INR), along with the Review Count, which indicates the number of customer reviews received. Additionally, it records the Rating, which represents the average user rating (on a scale of 1 to 5), and a Customer Review column that contains sample user feedback. This combination of structured data (such as price, rating, and review count) and unstructured data (such as textual descriptions and reviews) makes the dataset valuable for analyzing product trends and developing recommendation systems. The dataset provides insights into customer preferences and product performance across different online shopping platforms, helping users make informed purchasing decisions.

1	Category	Channel	Product I	Title	Description	Price	Review C	Rating	Customer Review
2	Home Appliances	Ajio	24daa2dc-0	Refrigerator	Energy-efficient fridge with smart cooling technology.	4870.77	9549	3.9	Energy efficient and works silently.
3	Beauty	Nykaa	8d87f2a5-2	Lipstick	Long-lasting matte lipstick with vibrant shades.	12752.65	2786	4.2	Works well, but packaging could be improved.
4	Beauty	Nykaa	87589356-5	Face Cream	Moisturizing face cream with SPF 30 protection.	3262.38	6181	3.5	Amazing product! Keeps my skin hydrated.
5	Home Appliances	Flipkart	cf2b7740-3	Refrigerator	Energy-efficient fridge with smart cooling technology.	24419.46	9565	4.2	Easy to use and has great features.
6	Clothing	Meesho	2cdd7513-b	Men's T-Shirt	100% cotton, comfortable fit, available in multiple colors.	12566.65	6045	4.4	The fabric is soft and comfortable to wear.
7	Home Appliances	Myntra	833acd34-8	Refrigerator	Energy-efficient fridge with smart cooling technology.	11227.45	8479	2.6	Good quality but expensive.
8	Home Appliances	Amazon	c663d58b-a	Refrigerator	Energy-efficient fridge with smart cooling technology.	341.2	6381	3.5	Easy to use and has great features.
9	Beauty	Amazon	4381c264-d	Lipstick	Long-lasting matte lipstick with vibrant shades.	15628.62	2009	2.9	Works well, but packaging could be improved.
10	Beauty	Amazon	bd4ad987-b	Lipstick	Long-lasting matte lipstick with vibrant shades.	23175.06	9833	4.3	Amazing product! Keeps my skin hydrated.
11	Beauty	Amazon	ec7cf677-et	Face Cream	Moisturizing face cream with SPF 30 protection.	16804.22	9428	3.2	Works well, but packaging could be improved.
12	Footwear	Myntra	235f9def-c5	Running Shoes	Lightweight and comfortable running shoes.	6191.96	5780	3.4	The sole is too hard, not recommended.
13	Beauty	Nykaa	388a97f9-7	Lipstick	Long-lasting matte lipstick with vibrant shades.	22518.96	3884	4.7	Long-lasting and has a great fragrance.
14	Electronics	Meesho	8c67dba6-2	Smartphone	Latest model with high-speed processor and long battery life.	12328.1	6331	4.5	The camera quality is amazing, but the battery drains fas
15	Home Appliances	Meesho	d63955a7-3	Refrigerator	Energy-efficient fridge with smart cooling technology.	1857.84	7904	3.5	Not as powerful as expected, takes longer to cook.
16	Home Appliances	Nykaa	35622363-6	Refrigerator	Energy-efficient fridge with smart cooling technology.	22781.18	6624	3.2	Good quality but expensive.
17	Clothing	Myntra	61466e36-8	Men's T-Shirt	100% cotton, comfortable fit, available in multiple colors.	7530.42	4885	3.6	Good quality but overpriced.
18	Clothing	Flipkart	cbfc0271-2t	Women's Dress	Elegant evening dress with a stylish design.	12790.83	2474	4	Good quality but overpriced.
19	Home Appliances	Flipkart	3f80ec33-a	Refrigerator	Energy-efficient fridge with smart cooling technology.	3505.62	4457	4.4	Good quality but expensive.
20	Footwear	Myntra	a2847ad2-9	Running Shoes	Lightweight and comfortable running shoes.	26889.39	672	4.5	Perfect for running, great grip and durability.
21	Beauty	Myntra	9d76a7f3-cl	Lipstick	Long-lasting matte lipstick with vibrant shades.	20336.5	4818	2.5	Not worth the hype, very average.
22	Home Appliances	Ajio	b51846ec-2	Refrigerator	Energy-efficient fridge with smart cooling technology.	21027.61	9384	4.8	Energy efficient and works silently.
23	Home Appliances	Flipkart	db525f24-9	Refrigerator	Energy-efficient fridge with smart cooling technology.	26301.62	6051	2.8	Energy efficient and works silently.
24	Home Appliances	Myntra	85db9ee0-0	Refrigerator	Energy-efficient fridge with smart cooling technology.	24792.59	5898	4.2	Easy to use and has great features.
25	Beauty	Flipkart	25117dc5-8	Lipstick	Long-lasting matte lipstick with vibrant shades.	7919.25	9272	4.3	Amazing product! Keeps my skin hydrated.
26	Clothing	Nykaa	a68785d5-d	Men's T-Shirt	100% cotton, comfortable fit, available in multiple colors.	21672.76	9893	3	The fabric is soft and comfortable to wear.
27	Home Appliances	Nykaa	4b1da4e0-9	Refrigerator	Energy-efficient fridge with smart cooling technology.	3310.19	411	4.7	Easy to use and has great features.
28	Beauty	Flipkart	16386dfe-1	Face Cream	Moisturizing face cream with SPF 30 protection.	29010.73	1855	3.5	Long-lasting and has a great fragrance.
29	Electronics	Meesho	3ea9083d-9	Laptop	Powerful laptop with an Intel i7 processor and 16GB RAM.	15435.94	1644	4.7	Decent product, but expected better display quality.

updated_merged_product_dataset

Fig 2: Ecommerce Dataset

5. Data Preprocessing

Data preprocessing was performed to enhance the quality and reliability of the dataset. Key steps included:

- **Handling Missing Values:** Missing data in numerical attributes (e.g., price, rating) was imputed using mean values, while missing textual data was handled by removing incomplete entries.
- **Categorical Encoding:** Label encoding was used to convert categorical variables such as product categories and channels into numerical representations.
- **Text Processing:** Customer reviews were tokenized and transformed into padded sequences using the Tokenizer and pad_sequences functions from TensorFlow.
- **Feature Scaling:** StandardScaler was applied to normalize numerical attributes such as price and ratings to improve model efficiency.

Additional steps involved stemming and lemmatization of textual data to enhance the representation of words in customer reviews. Stopwords and special characters were removed to improve the quality of text inputs. To optimize the processing of large text data, word embeddings such as Word2Vec and FastText were explored to create dense vector representations of words.

The code snipp*et al*lows users to interactively choose a category from a dataset and subsequently filter the data based on that selection. Initially, it gathers the unique values from the "Category" column by utilizing df['Category'].unique(), which generates a list of distinct product categories. These categories are then presented as a numbered list through a for loop, enabling users to clearly view their options. The script subsequently requests the user to enter a number that corresponds to their preferred category. After the user makes a selection, the script retrieves the chosen category by indexing into the unique_categories list. The selected category is then displayed to confirm the user's choice.

```
Choose a category:
1. Home Appliances
2. Beauty
3. Clothing
4. Footwear
5. Electronics
6. Mobile Phones
7. Chairs
8. Non-Fiction Book
9. Bats
10. Headphones
11. Fiction Book
12. Tables
Enter the number corresponding to your category:
```

Fig 3: User Input Category

6. Visualisation

Data visualization refers to the graphical representation of data, which facilitates the identification of patterns, trends, and insights. This process aids in comprehending intricate datasets by converting raw numerical information into visual formats such as charts, graphs, and maps. Common methods of visualization include bar charts, line graphs, scatter plots, histograms, and heatmaps.

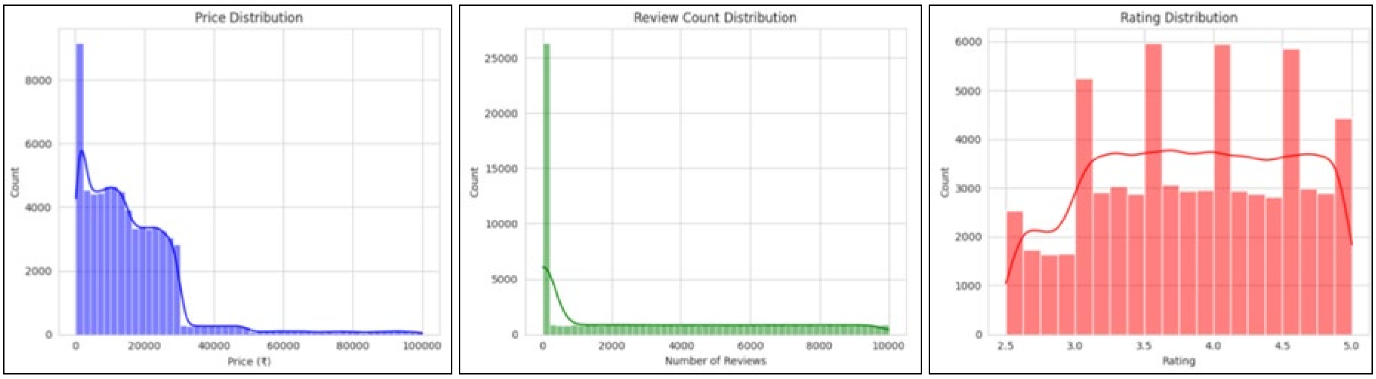

Fig 4: Histogram Chart

Each subplot corresponds to the distribution of a distinct feature:

- The first histogram (axes[0]) illustrates the distribution of product prices, employing sns.histplot() with 50 bins for detailed granularity and enabling KDE. This histogram is rendered in blue and is appropriately labeled.
- The second histogram (axes[1]) represents the distribution of review counts, utilizing analogous settings with 50 bins and KDE, and is displayed in green.
- The third histogram (axes[2]) showcases the distribution of product ratings, configured with 20 bins, KDE, and a red color scheme.

7. Model Architecture

A deep learning model was implemented using TensorFlow and Keras to recommend the most relevant products. The model architecture included:

- **Input Layer:** Accepting structured features (price, rating) and unstructured features (text reviews).
- **Hidden Layers:**
  Fully connected layers with ReLU activation.
  Batch Normalization to stabilize learning.
  Dropout layers (40%) to prevent overfitting.
- **Output Layer:** A softmax activation function was used to classify products into different categories based on learned patterns.

Hyperparameter tuning was conducted using Grid Search and Bayesian Optimization to determine optimal learning rates, batch sizes, and activation functions.

8. Training and Testing

The dataset was split into 80% training and 20% testing using train_test_split. The model was trained using:

- **Loss Function:** Sparse Categorical Cross-Entropy.
- **Optimizer:** Adam optimizer with a learning rate of 0.0005.
- **Evaluation Metrics:** Accuracy, precision, recall, F1-score, and R^2 score.
- **Batch Size:** 64.
- **Epochs:** 50.

Cross-validation was employed to prevent overfitting, and early stopping was utilized to terminate training when no further improvements were observed. Data augmentation techniques were applied to text reviews to improve model generalization.

Table 1: Training & Testing Evaluation

Metrics	Training	Testing
R2 Score	99	97
Accuracy	97.71	80.31

9. Recommendation Mechanism

The trained model was used to predict the most suitable product based on user-defined inputs. The recommendation process involved:

- Predicting the probability of each product category.
- Choosing the category that exhibits the greatest likelihood.
- Filtering the dataset to display the most relevant product for the user.

Table 2: Product Recommendation

Category	Channel	Product ID	Title	Price	Rating
Home Appliances	Flipkart	0edb8d3f-278c-4431-888e-9de5875987b6	Refrigerator	15009.36	4.9
Beauty	Ajio	46206dc5-3a2e-429f-90b5-b67dc9ae8dea	Face Cream	7985.71	3.5
Clothing	Meesho	17d05f4b-b43c-455d-b09a-d2a25c969d79	Women's Dress	7086.06	4.6
Footwear	Ajio	0af9b302-03a5-480e-b737-1f14c1ad1d34	Running Shoes	24139.42	3
Electronics	Ajio	b22c8665-df6c-490a-bfc5-7520b95c5cd4	Smartphone	19620.67	3.4
Mobile Phones	Flipkart	fe5d4c27-ebd4-4e3c-9b1a-f783bbe5de2b	Smartphone with 6.5-inch Display	25999.47	4.8
Chairs	Amazon	de6ccd16-c95d-4e14-bc46-fa68ba2de633	Ergonomic Office Chair	4517.59	3.9
Non-Fiction Book	Amazon	b05065cb-970a-4c49-9c02-437b0d3b40ab	Self-Help Book	1658.6	3.2
Bats	Amazon	bdfb4233-be3a-44a1-8e01-b29c264c9b7d	Cricket Bat	4463.65	4.9
Headphones	Flipkart	ed6f1799-03cb-4e39-b1cf-aba46a0a6a4d	Wireless Bluetooth Headphones	8377.85	5
Fiction Book	Nykaa	f53a88b2-a5bc-4968-8e3c-174f4b7a603e	Mystery Novel	302.02	3.9
Tables	Flipkart	e132cbe9-a119-44c6-aa47-27021b13aaa0	Wooden Dining Table	30410.67	3.3

10. Model Evaluation

The performance of the model was assessed using:

- **Accuracy:** 97.71% on training data, 80.31% on testing data.
- **R² Score:** 99% (training), 97% (testing), indicating strong predictive capability.
- **Loss Minimization:** Continuous monitoring of loss reduction across epochs ensured model stability.

Model robustness was further validated using external datasets to assess generalization across different product categories. A/B testing was conducted to compare model-generated recommendations with actual user preferences.

Table 3: Recommendation Evaluation

Metrics	Recommendation
R2 Score	98
Accuracy	96.71

11. Conclusion

The proposed deep learning-based product recommendation system effectively integrates structured and unstructured data to enhance decision-making in e-commerce. The use of advanced neural networks, feature encoding, and natural language processing significantly improves accuracy and personalization in recommendations. Future work will focus on refining model interpretability and incorporating hybrid recommendation techniques to further improve prediction accuracy. The integration of sentiment analysis from customer reviews is a potential avenue for enhancing recommendation quality. Additionally, explainable AI (XAI) techniques can be explored to provide transparency in recommendation decisions, helping users understand why specific products are recommended. The deployment of the model into a real-time recommendation engine and its integration with e-commerce platforms will be considered for future enhancements.

References

1. Caiwen Li, Iskandar Ishak, Hamidah Ibrahim, Maslina Zolkepli. Deep Learning-Based Recommendation System: Systematic Review and Classification. In Proceedings of the *IEEE Journals & Magazine,* 2023.
2. Shanthi S, Nirmaladevi K, Mathiarasi N & Tharani AK. Product Recommendation System using Deep Learning-based Recurrent Neural Network. In Proceedings of the 2023 *International Conference on Computer Communication and Informatics (ICCCI). IEEE,* 2023.
3. Zhang S. Study on Deep Learning-Based Personalised Product Recommendation Model for Autonomous Question-and-Answer Robot, 2025, 59–72. Springer.
4. Tayade A & Sejpal V. Deep Learning-Based Product Recommendation System and Its Applications, 2021.
5. Xu J, He X & Li H. Deep Learning for Matching in Search and Recommendation. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18), 2018, 1365–1368.
6. Yulianton H & Santi RCN. Product Matching using Sentence-BERT: A Deep Learning Approach to E-Commerce Product Deduplication. *Engineering and Technology Journal*. 2024; 9(12):5659–5664.
7. Foxcroft J, Chen T, Padmanabhan K, Keng B & Antonie L. Product Matching Lessons and Recommendations from a Real World Application. In Proceedings of the 34th Canadian Conference on Artificial Intelligence, 2021, 1–8.
8. Cheng Y & Li X. Product Matching in eCommerce. Walmart Global Tech Blog, 2017. DOI: https://medium.com/walmartglobaltech/product-matching-in-ecommerce-4f19b6aebaca
9. Met I, Erkoc A, Seker SE, Erturk MA & Ulug B. Product Recommendation System with Machine Learning Algorithms for SME Banking. *International Journal of Intelligent Systems*, 2024; (1):5585575.
10. Gai K, Qiu M & Zhao H. Privacy-preserving data mining in cloud computing. *Journal of Innovation & Knowledge*. 2024; 9(1):100097.