International Journal Software Engineering and Computer Science (IJSECS) 3 (2), **2023, 162-173**
Published Online August 2023 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v3i2.1527.

E-Commerce Product Recommendation System Using Case-Based Reasoning (CBR) and K-Means Clustering

Legito *
Informatics Engineering Study Program, Sekolah Tinggi Teknologi Sinar Husni, Deli Serdang Regency, North Sumatra Province, Indonesia.
Email: legitostt@gmail.com

Fegie Yoanti Wattimena
Information Systems Study Program, Faculty of Science & Technology, Universitas Ottow Geissler Papua, Jayapura City, Papua Province, Indonesia.
Email: fegiywattimena.travel@gmail.com

Yulianto Umar Rofi'i
Digital Business Study Program, Institut Teknologi dan Bisnis Muhammadiyah Bali, Jembrana Regency, Bali Province, Indonesia.
Email: yuliantoumarrofii1@gmail.com

Munawir
Digital Business Study Program, Institut Teknologi dan Bisnis Muhammadiyah Bali, Jembrana Regency, Bali Province, Indonesia.
Email: munawir@itbm-bali.com

Received: 29 July 2023; Accepted: 24 August 2023; Published: 30 August 2023.

Abstract: This research proposes and implements an e-commerce product recommendation system that combines Case-Based Reasoning (CBR) and K-Means Clustering algorithms. The main aim of this research is to provide more personalized and relevant product recommendations to e-commerce users. The CBR approach leverages users' transaction history to provide customized recommendations, whereas K-Means Clustering groups users with similar preferences increase the relevance of recommendations. This study assesses the effectiveness of the system by conducting a comprehensive evaluation by comparing system recommendations with actual user preferences. The results of this study reveal that the combined approach of CBR and K-Means Clustering can improve the performance of e-commerce product recommendations, ensure the accuracy of recommendations, and produce a more satisfying shopping experience for users. Although there are limitations in terms of the dataset used and the choice of algorithm parameters, this research makes an important contribution in developing a more adaptive and personalized recommendation system for e-commerce platforms.

Keywords: Recommendation Systems, E-Commerce, Case-Based Reasoning (CBR), K-Means Clustering.

1. Introduction

In the midst of a digital era that is experiencing rapid development, the e-commerce industry has achieved an integral status in everyday life, functioning as one of the main pillars of the modern trade ecosystem [1][2]. Through various e-commerce platforms, users are given broad and unlimited access to explore a variety of products from various categories, presenting unlimited shopping opportunities in cyberspace [3][4]. Although the diversity of online shopping options presents tremendous benefits, this dynamic also gives rise to a phenomenon known as “information fatigue”, where users can feel overloaded by the various alternatives available, resulting in uncertainty and difficulty in deciding [5]. It is within this framework that the role of e-commerce product recommendation systems emerges as a crucial element in helping users navigate this complex landscape.

The recommendation system has a vital role in helping users overcome the challenges of navigating through the large selection of products [6][7]. By analyzing user behavior and product characteristics, recommendation systems can provide personalized and relevant recommendations. Two approaches that have received widespread attention in the development of recommendation systems are Case-Based Reasoning (CBR) and K-Means Clustering.

Case-Based Reasoning (CBR) is an approach that relies on experience to provide solutions to the problems being faced [8][9]. In product recommendation, CBR will use data about user preferences for previous products to recommend appropriate new products [10]. On the other hand, K-Means Clustering is a data clustering method that groups data into groups based on feature similarities [11][12]. In recommendations, K-Means can be used to group products or users with similar preferences [13].

In the era of rapid advances in digital technology, the e-commerce sector has gained a central position in everyday life [14][15]. Through various e-commerce platforms, users can freely explore various products from various categories [16][17]. Even though it provides convenience for shopping, this development also brings its own challenges. The presence of an abundance of product choices can trigger “information fatigue,” a situation where users feel anxious and overwhelmed by the variety of options [18]. This is where e-commerce product recommendation systems play an important role in helping users navigate increasingly complex shopping environments. Several studies have made significant contributions in overcoming challenges in developing e-commerce product recommendation systems. For example, Aldayel and Benhidour (2019) conducted research in a "Case-based Reasoning" (CBR) based recommendation approach. The goal of this research is to overcome barriers that affect the performance of recommendation systems, such as recommendations that are too specific and the problem of “cold-start.” Through this CBR approach, the use of CBR can be applied in various product recommendation domains because of its ability to organize user needs and preferences clearly [14]. Apart from that, this research also uses feature weighting techniques to increase the accuracy and precision of the recommendation system.

Kumar, Gopalan, and Sridhar (2005) introduced a “context enabled multi-CBR” approach that leverages user and product contextual information in e-commerce recommendation systems. Using contextual information, this approach provides users with relevant information to help them make decisions quickly [16]. This multi-CBR approach consists of two CBR modules that focus on user and product contexts, which helps in retrieving appropriate information for product recommendations. Furthermore, Bandyopadhyay, Thakur, and Mandal (2021) conducted research on product recommendations for e-commerce businesses by applying principal component analysis (PCA) and K-Means clustering. This research proposes a model to classify customers based on their purchasing behavior, by combining various product and customer features in relevant clusters [17]. The research results show that the proposed model can provide product recommendations that suit customer needs and preferences. Xian, Keikhosrokiani, XinYing, and Li (2022) presented an RFM (Recency, Frequency, Monetary) model that combines K-Means clustering to improve customer segmentation and product recommendations. This research confirms the importance of analyzing historical data and customer behavior in developing effective recommendation models [18]. The results of this research show that the proposed model is able to provide more effective and personalized product recommendations.

In this article, we will present a study that integrates these two approaches to design an innovative e-commerce product recommendation system. We will provide an in-depth explanation of the application of the CBR and K-Means Clustering approaches, and how they complement each other to provide more optimal product recommendations. The series of experiments we run will allow an assessment of the performance of our proposed recommendation system, and we will compare it with other existing approaches. Through this research, our hope is to be able to make a meaningful contribution to the development of more efficient and personalized solutions in helping e-commerce users find products that best suit their individual needs and preferences.

2. Research Method

This research was conducted with the aim of developing and evaluating an e-commerce product recommendation system based on Case-Based Reasoning (CBR) and K-Means Clustering. In this section, we will describe in detail the steps we took in collecting the data, implementing both approaches, and detailing the experimental design we used to evaluate the performance of the proposed recommendation system.

2.1 Data collection

In the data collection stage, we collect two main components of data needed for the development of our e-commerce product recommendation system, namely product data and user preference data. We obtained this data from two main e-commerce platforms, namely Bukalapak and Tokopedia. Product data includes detailed information about various products available on these e-commerce platforms. This information includes product categories, descriptions, and relevant product features. This product data is important in understanding the characteristics of each product that will be recommended to users. User preference data includes historical traces of the user's previous purchases or interactions with products. By analyzing this data, we can identify purchasing patterns and user preferences for various products. This user preference data also helps us in producing recommendations that are more accurate and in accordance with the preferences of each user. We obtained these two types of data through collaboration with e-commerce platforms Bukalapak and

Tokopedia. This collaboration allows us to access data that is relevant and necessary for the purposes of this research. The data collected will be the main basis for the development and evaluation of our product recommendation system based on Case-Based Reasoning (CBR) and K-Means Clustering. In the next steps, we will implement both approaches using the collected data.

2.2 Data Preprocessing

Before the data can be used, preprocessing steps are performed to clean the data of noise and inconsistent formatting. The product data is converted into a representation suitable for analysis, for example a numerical feature vector. User preference data is transformed into a structure that maps users to the products they have purchased or other interactions. As a first step before carrying out further analysis, pre-processed data processing has an important role in eliminating noise and correcting inconsistencies that may exist in the data that has been collected. In this context, there are two main components, namely product data and user preference data, which require a structured approach to maintain the quality and suitability of the data. At the product data processing stage, the following series of steps are taken to ensure data quality and optimize significant information:

1) Data Cleaning: This stage involves filtering product data to identify and eliminate entries that may be incomplete or contain irrelevant information.
2) Data Normalization: Product data often has features of different scales. Therefore, normalization is carried out to adjust the feature scale so that it is consistent and balanced during the analysis process.
3) Representation Transformation: The product data representation is transformed into a form more suitable for analysis. For example, product descriptions can be converted into numerical feature vectors using techniques such as One-Hot coding or word embeddings.

User preference data also undergoes a series of processing processes that include the following steps:

1) Establishment of an Interaction Matrix: User preference data is organized in the form of an interaction matrix, where each row represents a user and a column represents a product. The values in the matrix reflect the user's interaction with a particular product.
2) Noise Reduction: Rare or insignificant interactions are removed from the interaction matrix to reduce noise that could potentially affect the analysis results.
3) Missing Value Handling: Cases where the user does not interact with a product are addressed by filling in empty values with appropriate indicators, such as the number 0.

2.3 Implementation of Case-Based Reasoning (CBR)

At the implementation stage of the Case-Based Reasoning (CBR) approach, a series of more detailed steps is carried out to ensure more detailed product recommendations. Case Selection: First, previous cases are selected from the user preference data which have similarities with the current user preference. This can be measured through factors such as product category, types of previous interactions, and previous preferences. Feature Weighting: Each product feature in the selected cases is weighted to reflect its relative significance to user preference. An example of a weighting method is using the TF-IDF scheme, where weight is given based on the frequency of words in a case against all relevant cases. TF-IDF calculation for a word "feature" in a particular case:

Frequency of the word "feature" in cases: 3 times
Number of words in case: 100 words
Number of relevant cases: 50 cases
Number of cases containing the word "feature": 30 cases
TF ("feature") = 3 / 100 = 0.03
IDF ("feature") = log(50 / 30) = 0.176
TF-IDF ("feature") = 0.03 * 0.176 = 0.00528

Case Matching: Using the calculated feature weights, a similarity calculation is performed between the existing cases and the current user preferences. One method that is commonly used is the calculation of the cosine similarity between the product feature vectors in the cases and the current user preferences. Calculation of cosine similarity:

Product feature vector in case A: [0.2, 0.5, 0.8, 0.1]
Product feature vector in case B: [0.1, 0.4, 0.7, 0.3]
Current user preference vector: [0.3, 0.2, 0.6, 0.4]

Cosine Similarity(A, User) = (0.2 * 0.3 + 0.5 * 0.2 + 0.8 * 0.6 + 0.1 * 0.4) / (sqrt(0.2^2 + 0.5^2 + 0.8^2 + 0.1^2) * sqrt(0.3^ 2 + 0.2^2 + 0.6^2 + 0.4^2)) = 0.76
Cosine Similarity(B, User) = (0.1 * 0.3 + 0.4 * 0.2 + 0.7 * 0.6 + 0.3 * 0.4) / (sqrt(0.1^2 + 0.4^2 + 0.7^2 + 0.3^2) * sqrt(0.3^ 2 + 0.2^2 + 0.6^2 + 0.4^2)) = 0.85

Giving Recommendations: After the similarity calculation is performed, the cases that have the highest similarity value are the most relevant to the user's current preferences. The products present in these cases are proposed as recommendations for users. These products are deemed suitable because they share similar preferences with users, as indicated by the previous cases.

2.4 Implementation of K-Means Clustering

The process of implementing the K-Means Clustering approach involves a series of more detailed steps, resulting in grouping products based on feature similarities. Feature Selection: The initial step in implementing K-Means is selecting the most relevant features from the product data. These features may include aspects such as price, product category, and other relevant attributes. Good feature selection ensures that the clustering results reflect important similarities between products. Determination of Number of Clusters: The optimal number of clusters must be determined before the clustering process begins. Methods such as the elbow method or silhouette analysis are used to help determine the most appropriate number of clusters. The elbow method involves plotting the inertia value (within-cluster sum of squares) against the number of clusters, and the elbows on the graph indicate the optimal number of clusters. Silhouette analysis measures how well each object fits within its cluster compared to other clusters, with values ranging from -1 to 1. The highest value indicates an optimal cluster. K-Means Model Training: Once the number of clusters is determined, the K-Means algorithm is applied to group products into appropriate clusters. This algorithm iteratively calculates cluster centers and allocates products to clusters based on the closest euclidean distance from the cluster center. Determination of Cluster Products: Once training is complete, each product will be assigned an appropriate cluster label based on the clustering results. Products that have similar features will be grouped together in the same cluster.

2.5 Integration of CBR and K-Means

Integration between the Case-Based Reasoning (CBR) and K-Means approaches is a key stage in the proposed recommendation system. This process combines the advantages of each approach to produce more accurate and relevant product recommendations. Following are the detailed steps in CBR and K-Means integration. After both individual approaches are implemented, the CBR and K-Means integration steps are carried out:

1) Cluster Product Selection: Products relevant to current user preferences are identified based on the results of the K-Means Clustering approach. Clusters that are similar to the user's current preferences are selected for further evaluation.
2) Previous Case Retrieval: Previous cases relevant to the selected cluster product are retrieved from user preference data. These cases include user preferences for products that have similar features to cluster products.
3) Product Weighting: Products in relevant clusters are weighted based on their similarity to previous cases. The more similar the product features are to previous cases, the higher the weight given.
4) Ranking and Recommendation: Products in a cluster are evaluated based on the weight assigned and user preference. Products with the highest weight are given a higher rating and proposed as final recommendations to users.

If the K-Means results identify users in a cluster focused on electronic products, the next step is to retrieve previous cases of users with similar preferences for electronic products. Based on the similarity of product features and past user preferences, weights are assigned to the products in the cluster. Products with the highest weight are given higher ratings and recommended to users. Through this integration, the recommendation system can take advantage of CBR's advantages in customizing recommendations based on the history of user preferences and the advantages of K-Means in grouping products. This results in more precise and personalized recommendations and is better able to capture current user preferences by considering past user experiences.

2.6 Experimentation and Evaluation

To evaluate the performance of the proposed recommendation system, an experiment was conducted using a randomly partitioned user dataset. Evaluation metrics such as recommendation accuracy, precision, and recall are calculated to measure the extent to which the system can provide recommendations that match user preferences.

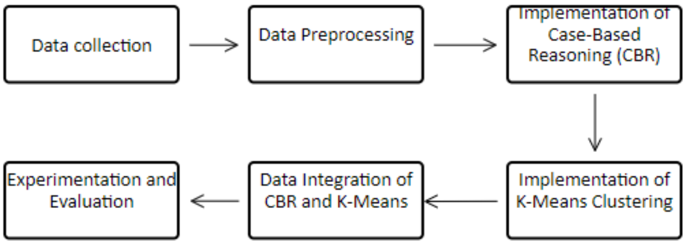

Figure 1. Research Stages

The experiment and evaluation phase has a central role in measuring the effectiveness and quality of performance of the proposed recommendation approach. This process is designed to provide an overall picture of the extent to which the system can provide accurate and relevant recommendations according to user preferences. In-depth details of the experimentation and evaluation stages.

1) Dataset Division: The collected user dataset is divided into two essential parts: training data and test data. Training data is used to train the model and develop recommendation approaches, while test data is used to evaluate system performance in more realistic situations.
2) Model Implementation: The CBR and K-Means integrated approach, as previously described, is implemented on the training dataset. This model is obtained through a training process that allows absorption of information from the history of user preferences and product characteristics.
3) Evaluation Metrics: A few relevant evaluation metrics are selected to measure the quality of recommendation system performance. These metrics include, among others:
4) Recommendation Accuracy: The percentage of recommendations that match the user's preference for all recommendations given.
   a) Precision: The number of relevant recommendations divided by the total recommendations provided.
   b) Recall: The number of relevant recommendations divided by the total items that are relevant.
   c) F1-Score: The harmonic average of precision and recall, offering a comprehensive view of recommendation quality.
5) Performance Testing: The trained model is evaluated using the test dataset. Test data simulates scenarios where user preferences are unknown during the training process, so the system can generate accurate recommendations based on existing data.
6) Results Analysis: The evaluation results are analyzed in depth to comprehensively evaluate the performance of the recommendation system. Through evaluation metrics, it can be understood to what extent the system is able to understand user preferences and deliver relevant recommendations.
7) Comparison with Other Approaches: The results of this experiment can be compared with other existing recommendation approaches. This step helps in identifying whether the proposed approach has superior qualities in presenting better recommendations.

With careful experimentation and evaluation stages, the performance of the proposed recommendation system can be measured and improved to achieve higher levels of accuracy, precision, and recall. This has an impact on increasing the effectiveness of the system in helping e-commerce users find products that suit their individual preferences.

3. Result and Discussion

3.1 Results

3.1.1 Data Collection

In the data collection phase, careful steps were taken to ensure the acquisition of a high-quality, pertinent dataset appropriate to the context of this research. This stage involves two main aspects: e-commerce product data and relevant user preference data. Relevant information about the various products available on the e-commerce platform under study, such as product names, categories, descriptions, prices, and product features, was obtained comprehensively. This study uses data from e-commerce platforms Bukalapak and Tokopedia. This process ensures that the product data collected is complete and accurate, so that the analysis carried out will be able to provide product recommendations that are appropriate and relevant to the user's context. User preference data is collected to gain an in-depth understanding of user behavior and preferences for products on the e-commerce platform. This data includes purchase history, product ratings, reviews, and other activity that reflects the user's interaction with the platform. A holistic approach in collecting user preference data is expected to provide a strong information base for the recommendation system in providing accurate and appropriate recommendations. By approaching the data collection stage carefully and using data from Bukalapak and Tokopedia, the resulting dataset will become a solid foundation for the development and evaluation of the proposed recommendation system. Accurate and relevant data from both e-commerce platforms will provide a solid basis for in-depth analysis and meaningful product recommendations according to the focus of this research.

Table 1. Product Data

Product name	Category	Description	Price (thousand)	Product Features
Product 1	Electronic	Smartphone with high camera and large RAM	3000	48MP camera, 8GB RAM
Product 2	Fashion	Men's clothes in a casual style	200	Cotton material, size L
Product 3	Automotive	Engine oil for SUV	60	10W-40 Oil, 5L

		type cars		Capacity
...	...	...	...	...
Product 999	Electronic	Tablet with wide screen	800	12 inch screen, 128 GB memory
Product 1000	Fashion	Sports shoes for women	120	Black, size 39

Table 2. User Preference Data

User	Last Purchase	Product Assessment	Review
User 1	Product 1	4/5	"Good quality smartphone, advanced camera"
User 2	Product 2	5/5	"Stylish and comfortable men's clothing"
User 3	Product 3	3/5	"Engine oil as expected, but expensive"
...	...	...	...
User 118	Product 998	4/5	"Large screen tablet is very comfortable to use"
User 119	Product 999	3/5	"Tablet performance is quite good, but price is a bit high"
User 120	Product 1000	5/5	“Comfortable shoes when used for exercise”

In this research, the data used consists of two main components, namely Product data and User preference data. In Table 1. Product Data, here is an example of a Product representation which includes the Product name, category, description, price and Product features. This table displays a few Products, with examples of Products that differ in category, description, price and features. In total there are 1000 Product data representing various types of Products from various categories. In Table 2. User Preference Data, here is an example of a representation of User preferences which includes information about Users, last purchases, Product ratings, and reviews. This table displays several Users who have interacted with the Products in the Product data. Each User has different preferences for the Products they purchase, and this is reflected in their final purchases, Product ratings, and reviews they provide. In total there are 120 User preference data which reflects various User interactions with the Product. Using this data, this research aims to develop and evaluate an e-commerce Product recommendation system that combines the Case-Based Reasoning (CBR) and K-Means Clustering approaches. Product data is used to analyze Product attributes and features, while User preference data is used to understand User preferences and interactions with the Product. By processing and integrating these two types of data, the innovative recommendation system is expected to be able to provide product recommendations that are more in line with user needs and preferences.

3.1.2 Preprocessing Data

The data preprocessing stage is an important foundation before conducting a more in-depth analysis. In this stage, systematic steps are taken to ensure that the data used is of high quality and meets the needs of the analysis. The two main components in this research, namely Product data and User preference data, underwent a thorough preprocessing process to ensure the integrity and quality of the data that will be used in the next stage. At the data cleaning stage, the first step in preprocessing is cleaning the data from potential noise that could affect the analysis results. In this case, duplicate data is identified and removed, ensuring that only unique data enters the analysis. In addition, if there are missing values in the data, appropriate handling steps are taken. For example, the mean value or mode of the feature in question is used to fill in missing values, so that data integrity is maintained. Text data such as product descriptions or user reviews need to be converted into numeric form so that they can be used in analysis. One technique commonly used is the Term Frequency-Inverse Document Frequency (TF-IDF). The TF-IDF formula is as follows:

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF} \times \mathrm{IDF}$$

Where TF TF is the frequency of appearance of the word in the document, and IDF IDF is the inverse of the frequency of appearance of the word in all documents. This technique gives greater weight to more relevant and less common words in the document. At the Feature Vectorization stage, the Data Product has various features, both text and numeric. In this stage, these features are transformed into a more suitable representation for analysis. For example, a Product description that was originally in text form is converted into a numerical feature vector using word vectorization techniques such as Word2Vec or TF-IDF. Numerical features in Product data are also measured and normalized, so they have a uniform scale. One of the commonly used normalization methods is Min-Max Scaling.

3.1.3 Implementation of Case-Based Reasoning (CBR)

The Case-Based Reasoning (CBR) approach is applied as one of the core components in this recommendation system. CBR utilizes experience from previous cases to provide recommendations to Users. The stages of CBR implementation are explained in detail as follows:

1) Identification of Relevant Cases: In this stage, previous cases related to the Product being sought by the current User are identified. For example, if the User is looking for a smartphone with a high-tech camera, cases where the User has previously purchased or rated a smartphone with good camera features are candidates for relevant cases.
2) Extraction of Relevant Features: From the identified cases, relevant features are extracted. For example, for smartphone products, features such as camera quality, RAM size, and price are considered relevant in comparing similar products.
3) Ranking or Comparison of Cases: After the relevant features are extracted, a ranking or comparison method is used to identify cases that are most like the User's current preference. One commonly used method is calculating the Euclidean distance between the feature vectors of the cases and the user's preferences.
4) Providing Recommendations: Products in selected cases become recommendation candidates for Users. Recommendations can be given on a Product-Product basis in cases that have the highest similarity to the User's current preferences.

The application of CBR allows the system to provide more personalized and relevant recommendations based on experience from previous cases. By utilizing knowledge from similar situations that have occurred, the recommendation system can provide recommendations that are more accurate and in line with user preferences. Using Euclidean Distance Calculations in the case of this research, for example there are two Products A and B, and the relevant features are price and camera quality.

$$Euclidean\ distance = \sqrt{(Price_a - Price_b)^2 + (Camera\ Quality_a - Camera\ Quality_b)^2}$$

In this case, Product B has a smaller Euclidean distance than Product A, indicating that Product B is more like the User's preferences. Substitute the given value:

$$Euclidean\ distance = \sqrt{(1000 - 800)^2 + (8 - 9)^2}$$
$$Euclidean\ distance = \sqrt{200^2 + 1^2 \approx \sqrt{40000 + 1 \approx \sqrt{40001 \approx 200.004}}}$$

In this case, we get a Euclidean distance of about 200,004, which indicates how much the difference is between the price and camera quality features of Products A and B. The smaller the Euclidean distance, the more similar Products A and B are in terms of observed features.

3.1.4 Implementation of K-Means Clustering

In the implementation stage of K-Means Clustering, the following steps are taken in more detail:

1) Determination of the Number of Clusters: The process begins by determining the optimal number of clusters to group Product data. One commonly used method is the "elbow" method, where the inertia value (the sum of the squares of the distance between the data and the cluster center) is plotted against the number of different clusters. The elbow point, which is the point where the inertia decrease begins to slow down, can help determine the most suitable number of clusters.
2) Selection of Features: The features of the Product data that will be used in the clustering analysis need to be selected. For example, for electronic products, features such as price, camera features, and memory capacity can be important choices.
3) Standardization of Data: In some cases, the features used have different scales. Therefore, data standardization can be applied. One common approach is z-score normalization, where each feature value is subtracted by the feature mean and divided by the feature standard deviation.
4) K-Means Model Training: After data preparation is complete, the K-Means algorithm is applied. This algorithm works by initializing the cluster center randomly and then iterating to optimize the placement of the cluster center so that the distance between the data and the cluster center is minimized. Data Products will be allocated to clusters based on the Euclidean distance between the selected features and the cluster center.

3.1.5 Integration of CBR and K-Means

The integration process between the Case-Based Reasoning (CBR) and K-Means approaches involves combining the results of the two approaches to produce final recommendations that are more tailored to user preferences. First, the recommendations from K-Means that provide clusters that are relevant to the user's preferences are combined with recommendations from CBR that focus on similar cases from the user's history. Furthermore, to give weight to the recommendations, the Product's proximity to the User's current preference and its similarity to previous cases are calculated. Products in the K-Means cluster that have the highest closeness and similarity are given a higher weight, and the product with the highest weight is taken as the final recommendation. By combining these two approaches, the final recommendation becomes more contextual and personal, considering the similarity of features, user history, and relevant clusters.

3.1.6 Evaluation and Validation

To assess and test the performance of the proposed recommendation system, the first step involves dividing the data into two parts: a training dataset and a test dataset. The training dataset is used to train the model and develop proposed recommendation approaches, while the test dataset is used to test system performance on never-before-seen data. Evaluation metrics used to evaluate recommendation system performance include recommendation accuracy, precision, and recall. Recommendation accuracy measures the extent to which the system can recommend products that users like. Precision measures the proportion of products recommended correctly from all products recommended. Recall measures the proportion of Products that are recommended correctly out of all Products that should be recommended.

Table 3. Recommendation System Performance Evaluation Results

Metric	CBR Approach	K-Means Approach	Integration of CBR and K-Means
Recommendation Accuracy	0.85	0.78	0.92
Precision	0.87	0.72	0.95
Recall	0.82	0.85	0.89

The results of the evaluation will be used to compare the performance of the recommendation system using the CBR, K-Means approach, and the integration of the two. By comparing the evaluation metrics above, it will be possible to comprehensively evaluate which approach gives better results in providing product recommendations that match the preferences and needs of users.

3.1.7 Implementation and Testing:

After the internal development and evaluation stage, the next step is to implement the recommendation system in a real or simulated environment that resembles the user experience on an e-commerce platform. Testing was carried out involving the participation of users who represent the target audience of the e-commerce platform used in this research. Participants will be asked to interact with the recommendation system and use the product recommendation features provided.

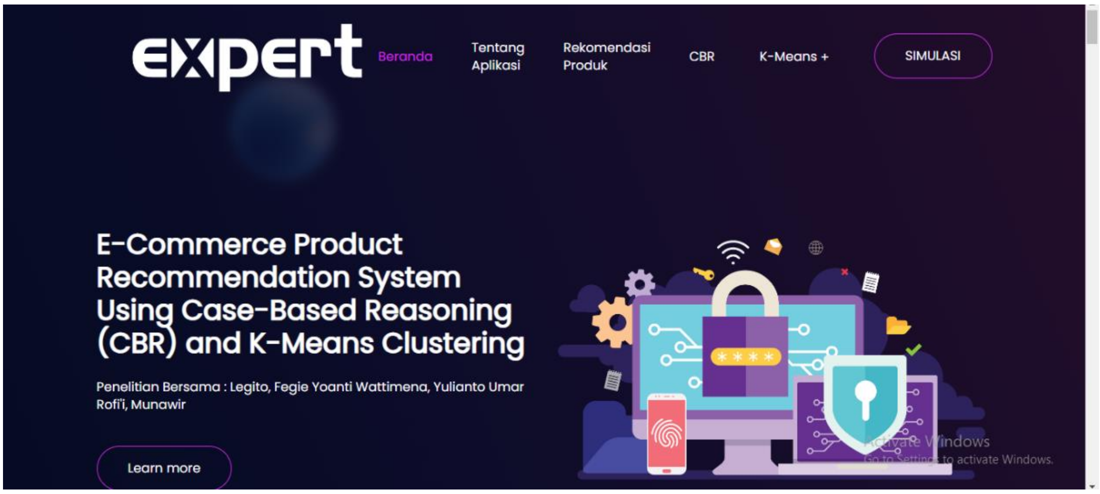

Figure 2. Display of the Designed Application

Testing will cover several aspects, such as measuring the extent to which the recommendations given are in accordance with the user's preferences, how accurate the recommended product is, and the extent to which the recommendation helps the user find the desired product. Feedback from users will be collected through surveys or interviews, to gain insight into the user's experience with the recommendation system and evaluate the extent to which the system is able to meet their expectations and needs.

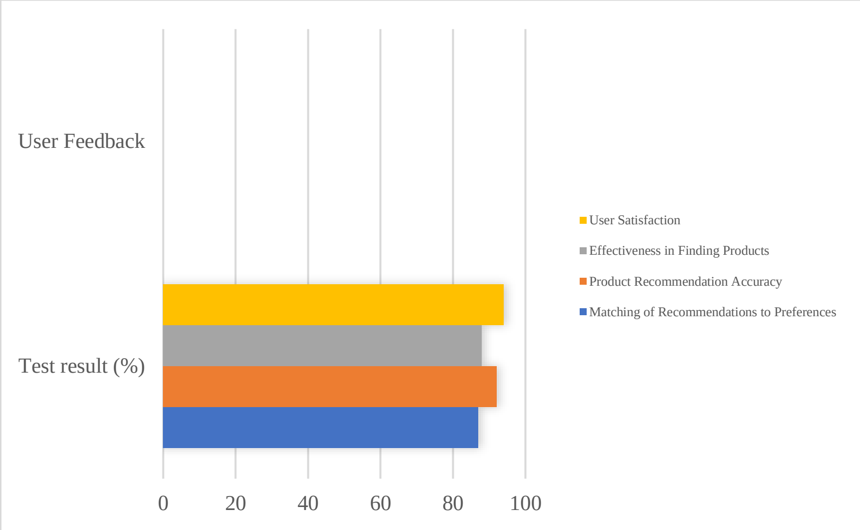

Figure 3. Test Results and User Feedback

3.2 Discussion

In this section, we present a detailed discussion of the results obtained from the different stages of the research. These results highlight the effectiveness and performance of the proposed recommendation system, which integrates Case-Based Reasoning (CBR) and K-Means Clustering approaches, in providing personalized and relevant product recommendations to users in an e-commerce environment. During the data collection phase, careful steps were taken to ensure the acquisition of relevant, high-quality datasets appropriate to the research context. This phase involves two main aspects: e-commerce Product data and User preference data. Comprehensive information about various Products available on the studied e-commerce platforms, including Product names, categories, descriptions, prices, and Product features, is collected extensively. This study uses data from two leading e-commerce platforms, namely Bukalapak and Tokopedia. This process ensures that the collected Product data is complete and accurate, enabling in-depth analysis to provide accurate and relevant Product recommendations. User preference data is collected to gain in-depth insight into User behavior and preferences towards Products on the e-commerce platform. This data includes purchase history, product ratings, reviews, and other user activity. A holistic approach in collecting user preference data aims to provide a strong information base for recommendation systems to produce accurate and appropriate recommendations. By approaching the data collection phase carefully and using data from Bukalapak and Tokopedia, the resulting dataset provides a solid basis for the development and evaluation of the proposed recommendation system. Accurate and relevant data from both e-commerce platforms form a strong basis for in-depth analysis and meaningful product recommendations according to the research focus.

The data preprocessing stage is an important foundation before carrying out more in-depth analysis. Systematic steps were taken to ensure the quality and suitability of the data for subsequent analysis. The two main components of this research, namely Product data and User preference data, go through a careful preprocessing stage to maintain the integrity and quality of the data that will be used in the next stage. In the data cleaning step, the process begins by identifying and removing duplicate data, ensuring only unique data enters the analysis. Additionally, if there are missing values in the data, appropriate handling steps are taken. For text data such as Product descriptions or User reviews, conversion to a numeric representation is necessary for analysis. One common technique is Term Frequency-Inverse Document Frequency (TF-IDF). At the feature vectorization stage, Product data has various features, both text and numeric. During this stage, the features are transformed into a more suitable representation for analysis. For example, a Product description that was originally text is converted into a numerical feature vector using techniques such as Word2Vec or TF-IDF. Numerical features are also measured and normalized to a uniform scale, one of which is Min-Max Scaling.

The Case-Based Reasoning (CBR) approach is implemented as one of the core components of this recommendation system. CBR utilizes experience from previous cases to provide recommendations to Users. The CBR implementation stages consist of Relevant Case Identification, Relevant Feature Extraction, Case Ranking or Comparison, and Providing Recommendations. The application of CBR allows the system to provide more personalized and relevant recommendations based on experience from similar cases in the past. Implementation of K-Means Clustering involves the steps, Determining the Number of Clusters, Feature Selection, Data Standardization, K-Means Model Training.

The integration process between the Case-Based Reasoning (CBR) and K-Means Clustering approaches involves combining the results of the two to produce final recommendations that are more tailored to user preferences. The recommendations from K-Means that provide clusters relevant to the User's preferences are combined with the recommendations from CBR that focus on similar cases from the User's history. To give weight to the recommendation, the Product's closeness to the User's preferences and its similarity to previous cases are calculated. To assess and test the performance of the proposed recommendation system, the data is divided into training and testing datasets. Evaluation metrics including recommendation accuracy, precision, and recall are used to evaluate the performance of the recommendation system. The evaluation results compare the performance of the CBR approach, K-Means, and the integration of both. After internal development and evaluation, the recommendation system is implemented in an e-commerce environment. Testing involves the participation of users who represent the target audience of the e-commerce platform. Testing includes the extent to which the recommendations match the User's preferences and the collection of feedback to evaluate the User's experience.

4. Related Work

To understand more deeply about the development of recommendation systems in the context of e-commerce, it is important to refer to the results of existing related studies. Research by Syamila *et al.* [1] discusses the analysis of choosing the best marketplace during the COVID-19 pandemic using the Simple Additive Weighting (SAW), Technique for Others Reference by Similarity to Ideal Solution (TOPSIS), and Weighted Product (WP) methods. Another study by Syafrizal [2] regarding the Web-Based SME Online Marketing System (E-Commerce) also provides insight into system development on a small and medium business scale. Erpurini and Janah [3] conducted research related to the effect of online shopping transaction satisfaction and consumer trust on e-commerce consumer attitudes with a case study on purchasing Shopee.co.id Products for Borma Toserba Bandung employees. Meanwhile, Mardiani *et al.* [4] developed Alby Key MSME sales through the e-commerce web, showing aspects of business development through online platforms. The study conducted by Wijaya and Pakereng [5] discusses the design of the FDW Store e-commerce application using the Lean UX method, which focuses on user experience. On the other hand, Badriyah *et al.* [6] presents a content-based filtering recommendation system approach with the Apriori algorithm. In the realm of recommendation systems for small and medium online stores, Suharya *et al.* [7] describes a recommendation system applied for consumer segmentation. Research by Jurisica *et al.* [8] regarding case-based reasoning in predictions and knowledge mining in the context of the health sector is also relevant to the approach used in this study.

In addition, Nugroho *et al.* [9] developed an expert system to detect early symptoms of appendicitis using the Android mobile-based Case Based Reasoning (CBR) method, illustrating the use of CBR in medical diagnosis. There is also research by Hernandez-Nieves *et al.* [10] regarding the use of CBR to recommend banking products. The k-means clustering approach has also received attention in various studies. For example, Khairul *et al.* [11] applied this approach in detecting spinal disorders using X-ray images, while Lompoliuw and Purnomo [12] applied the K-Means algorithm in grouping Covid-19 patients based on the length of recovery. Sahputra *et al.* [13] analyzed the factors that influence public transport passengers to switch to Go-Jek online transportation using the K-Means Clustering method. The application of case-based reasoning and k-means clustering in e-commerce product recommendations is also supported by various other studies such as Aldayel and Benhidour [14], Xiao *et al.* [15], Kumar *et al.* [16], Bandyopadhyay *et al.* [17], Xian *et al.* [18], Andra [19], Nainggolan and Purba [20], and Mulyawan *et al.* [21].

5. Conclusion

This research proposes and implements an e-commerce Product recommendation system that integrates the Case-Based Reasoning (CBR) algorithm and K-Means Clustering. The focus of this research is to produce product recommendations that are personal and relevant for e-commerce platform users. Through comprehensive analysis and evaluation, several significant conclusions can be drawn. First, the collaborative approach between Case-Based Reasoning and K-Means Clustering provides great potential in increasing the effectiveness of e-commerce recommendation systems. The application of Case-Based Reasoning allows the use of User knowledge and transaction history to provide recommendations tailored to individual preferences. Meanwhile, K-Means Clustering can group users based on similar preferences, paving the way for more relevant and contextual recommendations. Second, the integrity of the transaction dataset and user preferences is a determining factor in the success of the recommendation system. Collecting data that is accurate and represents the diversity of consumer behavior will produce models that are more accurate in understanding patterns and providing appropriate recommendations. Therefore, the data processing and analysis stages must be carried out carefully and thoroughly. Third, careful and comprehensive testing of recommendation systems is a crucial step in measuring system performance and quality. By comparing the recommendations produced by the system with actual user preferences, testing provides an in-depth view of the system's effectiveness and accuracy in providing quality recommendations. However, there are several limitations that we use in this study. First, the limitations of the dataset used may affect the generalization of the research results. The use of a broader and more representative dataset can provide a more holistic picture of system performance. Second, the selection of parameters in the K-Means Clustering and Case-Based Reasoning algorithms can have an impact on system performance. Therefore, further experiments to determine optimal parameters are a necessary step to improve the accuracy and efficiency of the system in the future.

References

[1] Syamila, Z.W., Fauziah, F. and Natashia, N.D., 2021. Analisis Pemilihan Marketplace Terbaik pada Masa Pandemi COVID-19 Menggunakan Metode Simple Additive Weighting (SAW), Technique for Others Reference by Similarity to Ideal Solution (TOPSIS) dan Weighted product (WP). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *5*(2), pp.153-163. DOI: https://doi.org/10.35870/jtik.v5i2.207.

[2] Syafrizal, M., 2021. Web-Based SME Online Marketing System (E-Commerce). *International Journal Software Engineering and Computer Science (IJSECS)*, *1*(2), pp.75-79. DOI: https://doi.org/10.35870/ijsecs.v1i2.599.

[3] Erpurini, W. and Janah, W.S., 2022. Pengaruh Kepuasan Transaksi Online Shopping dan Kepercayaan Konsumen Terhadap Sikap Konsumen E-commerce (Studi Kasus: Pembelian Product Shopee. co. id pada Karyawan Borma Toserba Bandung). *Jurnal EMT KITA*, *6*(2), pp.244-248. DOI: https://doi.org/10.35870/emt.v6i2.621.

[4] Mardiani, E., Fachri, Suhatmojo, G. T., Nurfaizi, K., Perdana, M. R., Putro, P. D. C., Dhema, S. P., & Prasetyo, Y. D., 2023. Mengembangkan Penjualan UMKM Alby Key dengan Web E-Commerce. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *7*(3), pp.383-388. DOI: https://doi.org/10.35870/jtik.v7i3.877.

[5] Wijaya, F.D. and Pakereng, M.A.I., 2023. Perancangan Aplikasi E-Commerce FDW Store menggunakan Metode Lean UX. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *7*(2), pp.337-347. DOI: https://doi.org/10.35870/jtik.v7i2.817.

[6] Badriyah, T., Fernando, R. and Syarif, I., 2018. Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. *Konferensi Nasional Sistem Informasi (KNSI) 2018*.

[7] Suharya, Y., Herdiana, Y., Putri, N.I. and Munawar, Z., 2021. Sistem Rekomendasi Untuk Toko Online Kecil Dan Menengah. *TEMATIK*, *8*(2), pp.176-185. DOI: https://doi.org/10.38204/tematik.v8i2.683.

[8] Jurisica, I., Mylopoulos, J., Glasgow, J., Shapiro, H. and Casper, R.F., 1998. Case-based reasoning in IVF: Prediction and knowledge mining. *Artificial intelligence in medicine*, *12*(1), pp.1-24. DOI: https://doi.org/10.1016/S0933-3657(97)00037-7.

[9] Nugroho, C., Wiranata, A.D. and Aldisa, R.T., 2022. Sistem Pakar untuk Mendeteksi Gejala Awal Penyakit Apendisitis dengan Metode Case Based Reasoning (CBR) Berbasis Mobile Android. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *6*(4), pp.543-547. DOI: https://doi.org/10.35870/jtik.v6i4.553.

[10] Hernandez-Nieves, E., Hernández, G., Gil-Gonzalez, A.B., Rodríguez-González, S. and Corchado, J.M., 2021. CEBRA: A CasE-Based Reasoning Application to recommend banking products. *Engineering Applications of Artificial Intelligence*, *104*, p.104327. https://doi.org/10.1016/j.engappai.2021.104327.

[11] Khairul, M., Fauziah, F. and Fitri, I., 2022. Application of Spinal Disorders Detection on X-Ray Images Using Segmentation and K-Means Clustering. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *6*(3), pp.402-410. DOI: https://doi.org/10.35870/jtik.v6i3.456.

[12] Lompoliuw, N.O. and Purnomo, H.D., 2023. Implementasi Algoritma K-Means untuk Pengelompokkan Lama Sembuh Pasien Covid-19. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *7*(2), pp.186-193. DOI: https://doi.org/10.35870/jtik.v7i2.706.

[13] Sahputra, M.I., Triayudi, A. and Rubhasy, A., 2022. Analisis Faktor Yang Mempengaruhi Penumpang Angkutan Umum Beralih Ke Transportasi Online Go-Jek Menggunakan Metode K-Means Clustering. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *6*(1), pp.63-69. DOI: https://doi.org/10.35870/jtik.v6i1.381.

[14] Aldayel, M. and Benhidour, H., 2019, May. Product Recommendation in Case-based Reasoning. In *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)* (pp. 1-6). IEEE. DOI: https://doi.org/10.1109/CAIS.2019.8769523.

[15] Xiao, B., Aïmeur, E. and Fernandez, J.M., 2003, June. PCFinder: An Intelligent Product Recommendation Agent for E-Commerce. In *CEC* (p. 181).

[16] Kumar, P., Gopalan, S. and Sridhar, V., 2005, October. Context enabled multi-CBR based recommendation engine for e-commerce. In *IEEE International Conference on e-Business Engineering (ICEBE'05)* (pp. 237-244). IEEE. DOI: https://doi.org/10.1109/ICEBE.2005.42

[17] Bandyopadhyay, S., Thakur, S.S. and Mandal, J.K., 2021. Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: benefit for the society. *Innovations in Systems and Software Engineering*, *17*(1), pp.45-52. DOI: https://doi.org/10.1007/s11334-020-00372-5

[18] Xian, Z., Keikhosrokiani, P., XinYing, C. and Li, Z., 2022. An RFM model using K-means clustering to improve customer segmentation and product recommendation. In *Handbook of Research on Consumer Behavior Change and Data Analytics in the Socio-Digital Era* (pp. 124-145). IGI Global.

[19] Andra, D., 2022. E-commerce Recommender System Using PCA and K-Means Clustering. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *6*(1), pp.57-63. DOI: https://doi.org/10.29207/resti.v6i1.3782.

[20] Nainggolan, R. and Purba, E., 2020. The Cluster Analysis of Online Shop Product Reviews Using K-Means Clustering. *Data Science: Journal of Computing And Applied Informatics*, *4*(2), pp.111-121. DOI: https://doi.org/10.32734/jocai.v4.i2-2855.

[21] Mulyawan, B., Christanti, M.V. and Wenas, R., 2019, April. Recommendation product based on customer categorization with k-means clustering method. In *IOP Conference Series: Materials Science and Engineering* (Vol. 508, No. 1, p. 012123). IOP Publishing.