IMPROVING RANKING QUALITY VIA CONSEQUENT SET QUARTILE PARTITIONING IN RULE-BASED RECOMMENDATION

Tubagus Mohammad Akhriza[1*], Khoerul Anwar[2], Weda Adistianaya Dewa[3]
[1,3]Dept. of Information System, STMIK PPKIA Pradnya Paramita
[2]Dept. of Information Technology, STMIK PPKIA Pradnya Paramita, Malang
Jl. Laksda Adi Sucipto 249A, Pandanwangi, Blimbing, Malang City, Indonesia
[1]akhriza@stimata.ac.id, [2]alqhoir@stimata.ac.id, [3]weda@stimata.ac.id

Abstract – Accurate ranking in session-based recommender systems remains a challenge, especially in dynamic domains such as e-commerce and digital news where ranking quality directly shapes user satisfaction and platform effectiveness. Neural models achieve high accuracy but incur heavy computational costs, while association rule (AR)-based methods are efficient yet struggle to prioritize relevant items when multiple candidates share similar weights. This study aims to improve the ranking quality of AR-based recommendation while retaining computational efficiency. A lightweight framework, ART-Q (Association Rule with Top-k Quartile Filtering), is proposed to extend traditional single top-k ranking into multiple top-k selection across quartiles. The approach follows three phases: (i) modeling—constructing an association rule dictionary from training data, maintaining antecedents as keys with consequent sets and rule weights as values; (ii) inference—partitioning each consequent set into four quartiles and selecting top-k items from each, thereby producing a multiple top-k recommendation that surfaces candidates across weight strata; and (iii) evaluation—benchmarking ART-Q against non-neural (AR, k-Nearest Neighbor variants) and neural baselines (recurrent, convolutional, and graph-based models) using standard ranking metrics. Experiments on two real-world datasets from e-commerce and e-news domains show that ART-Q consistently improves hit rate, reciprocal rank, and ranking stability while preserving computational efficiency. Nevertheless, reliance on static quartile partitioning and the limited dataset scope constrain generalizability. Future work may explore adaptive weighting strategies and broader validation to strengthen applicability across domains.

Keywords: Recommendation system, Association Rules, Session-based Recommendation, Ranking quality, Top-k recommendation

I. INTRODUCTION

Recommender systems (RS) play a critical role in digital platforms such as e-commerce and online news, where they personalize user experiences by suggesting items likely to be of interest [1], [2]. A particular challenge arises in anonymous platforms where explicit user preference data is unavailable. Session-based RS have been developed to address this issue by leveraging short-term user interactions within a session to predict the next items to be viewed in real time [3]–[5].

The urgency of improving these systems lies in the fact that ranking quality directly affects user engagement and satisfaction: when relevant items are placed too low in the recommendation list, the likelihood of user engagement diminishes, leading to suboptimal experiences and reduced platform effectiveness [6], [7].

Research on session-based RS has produced two broad families of methods. Neural models such as GRU4Rec [8], NextItNet [9] and SR-GNN [10] achieve high performance by modeling sequential or graph-structured dependencies, but their reliance on large datasets and significant computational resources limits their practicality in real-time deployments [4], [11]. Non-neural approaches, including Association Rule (AR) mining and k-Nearest Neighbor (KNN), remain attractive for their efficiency and interpretability. Prior studies have shown that with proper configuration, such methods can even rival or surpass neural models in certain conditions [4], [11].

In general, traditional methods select top-k recommendations from a single candidate set, typically those with the highest probabilities after observing the items in a session [12]–[14]. In AR-based methods, items are chosen from consequent sets that maximize support—the frequency of an itemset in the dataset—and confidence—the conditional probability of a consequent given the antecedent [15]. However, when multiple candidates share similar support–confidence values, ranking ambiguity arises: relevant items may be ranked lower, which reduces Hit Rate (HR), Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG). In addition, traditional AR methods in session-based recommendation typically rely only on the last item in a session as the antecedent reference for identifying consequent sets [11], [16], which restricts item coverage and may even fail to generate relevant recommendations.

To address these challenges, this study introduces Association Rule with Top-k Quartile Filtering (ART-Q). ART-Q reorganizes consequent sets into four quartiles (Q1–Q4) ordered by support–confidence weights and selects multiple top-k items across them. In addition to selecting candidates from each quartile, ART-Q utilizes all items in the session as antecedent references to generate multiple top-k recommendations per consequent set. By diversifying candidate coverage in this way, ART-Q aims to reduce ranking ambiguity and improve metrics such as HR, MRR and NDCG, without substantially increasing computational complexity.

The objective of this research is therefore to improve the ranking quality of AR-based session recommendation in dynamic domains while retaining computational efficiency. To this end, quartile-based filtering is introduced as a refinement strategy, its effectiveness is evaluated on two real-world datasets—e-commerce (YooChoose) and digital news (MPM)—and its performance is analyzed at both quartile and antecedent levels to provide deeper insights into how rule-based recommenders can be enhanced for session recommendation tasks.

II. RESEARCH METHODS

The general workflow of ART-Q is illustrated in Figure 1. Broadly, it is divided into three main phases: the Modeling Phase, the Inference Phase, and the Evaluation Phase. The details of each phase are provided in the following subsections. However, before that, the formalization of the ART-Q model is outlined first.

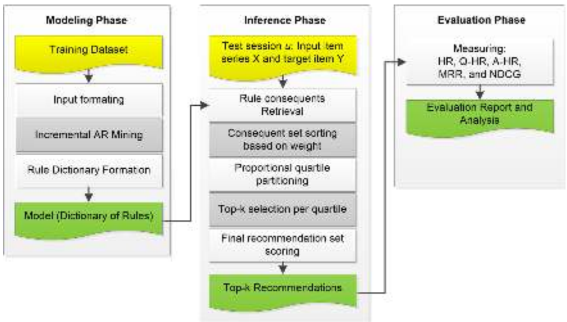

Figure 1. General Workflow of ART-Q method

3.1. Model Formalization

Given a session $u = \{x_1, x_2, \dots, x_i\}$ of clicked items, where each x_i is unique within the set, and let $Y = \{Y_1, Y_2, \dots\}$ denote the set of consequents set Y_i associated with each $x_i \in u$, following prior formulations of next-item recommendation [9], [17], [18].

ART-Q extends this formulation by aggregating rule-based predictions across all antecedents, so that the probability distribution accounts for the combined influence of every x_i and its consequents $y \in Y_i$ (the next-item candidates for session u). Each pair $x_i \rightarrow Y_i$ thus constitutes an association rule, and the overall probability distribution of recommendations is derived by combining these rules, as shown in Equation (1).

$$P(y|u) = \sum_{x_i \in u} \sum_{y \in Y_i} P(y|x_i) \times w(x_i, y) \tag{1}$$

Where $w(x_i, y) = conf(x_i, y) \times \sup(y)$ represents the weight of y relative to x_i. Prior literatures of general AR-based methods often employ the Lift metric to indicate the correlation strength between x and y [7], [19]–[22], which is normalizing the rule confidence by the support of y, or $Lift(xy) = \frac{conf(x,y)}{\sup(y)}$. However, ART-Q deliberately avoids using Lift because such normalization may unintentionally penalize the relevance of less frequent items. Additionally, this weighting approach mitigates uniformity in rule scoring, preventing excessive reliance on confidence alone. After sorting the items in Y_i by their highest weights, they are divided into four quartiles Q_1, Q_2, Q_3, Q_4, where each quartile covers a proportion p_i, ensuring that $\sum_{i=1\ldots4} p_i = 1$. Recommendations are generated by independently selecting top-k items from each quartile after ordering items within Q_i by their highest weights.

Let $Topk\text{-}Q_i$ represent the set of $y_j \in Q_i$ where $j \leq k$, with the highest weights in Q_i. The probability of selecting item y within individual quartile Q_i, is given in Equation (2):

$$P(y|u, Q_i) = \sum_{x_i \in u} P(y|x_i, Topk\text{-}Q_i) \times w(x_i, y) \tag{2}$$

The total probability across all quartiles is then computed in Equation (3):

$$P(y|u, Q) = \sum_{j=1\ldots4} P\big(y|u, Q_j\big) \tag{3}$$

Given that the recommendation space is constrained within $Topk\text{-}Q_i$, the probability for an individual antecedent x_i across all quartiles with respect to $Topk\text{-}Q_i$ is defined in Equation (4):

$$P(y|x_i) = \sum_{j=1\ldots4} P(y|x_i, Topk\text{-}Q_i) \times w(x_i, y) \tag{4}$$

Since multiple antecedents contribute to session-based recommendations, ART-Q computes the total session-level probability by aggregating over all antecedents. Thus, the final probability is derived in Equation (5):

$$P(y|u) = \sum_{x_i \in u} P(y|x_i) \tag{5}$$

3.2. Modeling Phase

The modeling phase of ART-Q constructs a rule dictionary from the training dataset T, capturing relationships between antecedents and their corresponding consequent sets. In this dictionary, antecedents serve as keys, while their consequent sets function as values, guiding the recommendation process. The rule construction and dictionary formation mechanism are outlined as the following steps. The first stage involves input formatting, where the session data in dataset T is preprocessed to ensure that each row corresponds to a unique SessionId, containing a sequence of ItemId entries ordered by their respective timestamps.

Subsequently, incremental AR mining is applied. For each SessionId, the items within that session are used to form 1-itemsets and 2-itemsets through permutation. A frequency counter is maintained to increment the occurrence of individual items and item pairs (x, y). After the counting step, association rules of length two are constructed from the mined 2-itemsets, representing simple item-to-item transitions within sessions. An advantage of this incremental mining approach is that it enables fast and continuous rule construction in streaming environments, where transaction data arrives session by session [19], [23]. This eliminates the need to wait for a full data accumulation window (for example one day or week) before updating the model. As a trade-off, the computed support and confidence values are approximations, yet they remain effective in capturing the relative strength and relevance of item associations.

In the final stage, a rule dictionary is formed. After processing all sessions in T, a dictionary of rules is constructed. Each key corresponds to an antecedent item x, and its value is a nested dictionary of consequent items y paired with their corresponding rule weights $w(x, y)$. Since the supports of x, y, and their co-occurrence (x, y) have been collected in step 2, computing this weight is straightforward. The final dictionary format is structured as $x_i\colon \{y_1\colon w_1, y_2\colon w_2, \dots\}$. Once all records are processed, the resulting rule dictionary fully encapsulates the learned model from the session data, serving as the foundation for the subsequent inference and recommendation phases.

3.3. Inference Phase

In real-world scenarios, the inference phase processes a session $u = \{x_1, \dots, x_{i+1}\}$ to generate multiple next-item predictions $x_{i+1}, x_{i+2}, \dots$. However, during the experiment, each session u in the test set is split into two parts: $u_x = \{x_1, \dots, x_i\}$ representing the observed antecedents, and $y = x_{i+1}$, serving as the ground-truth next item. The model aims to predict y within the top-k recommendations by following these steps.

The process begins with rule consequents retrieval. For each antecedent $x \in u_x$, retrieve its set of consequents Y from the rule dictionary, where each consequent $y \in Y$ is already associated with a rule weight $w(x, y)$. Next, the consequent set is sorted. All consequent items $y \in Y$ are sorted in descending order based on their rule weights $w(x, y)$, yielding a ranked list of candidate consequents for each antecedent.

Following the sorting stage, proportional quartile partitioning is performed. The ranked list of consequent items is partitioned into four quartiles: Q_1, Q_2, Q_3, Q_4 representing segments from the highest to the lowest rule weights, respectively. Each quartile is assigned a proportional weight $p_1 : p_2 : p_3 : p_4$, such that $\sum_{i=1..4} p_i = 1$. This quartile-based partitioning serves two primary purposes. First, it enables low-scoring but contextually relevant consequents, often found in Q_3 and Q_4. Second, it allows an item near the tail end of a higher quartile, such as Q_2, to be re-ranked into the head of a lower quartile Q_3 under top-k selection, thereby increasing its chances of inclusion in the final recommendation list.

Subsequently, top-k selection is applied to each quartile. From each quartile Q_i, a subset of k top-ranked items is selected, defined as $Topk\text{-}Q_i = \{y_j \in Q_i \mid j \leq k\}$, where y_j represents a candidate next-item selected within quartile Q_i. Through this approach, multiple top-k recommendations are generated across quartiles, ensuring that not only the most frequent items but also lower-weighted yet contextually relevant candidates are considered. Finally, the total recommendation probability is computed. The total probability across all quartiles is then computed in Equation (3), while total probability across all antecedents in session u is computed using (5).

3.4. Evaluation Phase

To assess the performance of baseline models, standard ranking metrics, including HR, MRR, and NDCG are employed [13], [24], [25]. HR@k quantifies the proportion of sessions where the ground-truth item y appears in the top-k recommendations [26], [27], as defined in Equation (6), which quantifies the proportion of sessions where the ground-truth item appears in the top-k recommendations.

$$HR@k = \frac{1}{|U|} \sum_{u \in U} 1 \left(I_u \in R_u^k \right) \tag{6}$$

MRR measures the average reciprocal rank of the first correctly predicted ground-truth item, ensuring that highly relevant items appear earlier in the recommendation list [11], [27], [28]. It is defined in Equation (7):

$$MRR@k = \frac{1}{|U|} \sum_{u \in U} \frac{1}{\mathrm{rank}_u} \tag{7}$$

where $rank_u$ represents the position of the first correctly recommended item in session u. A higher MRR score indicates improved ranking quality, with relevant items appearing closer to the top.

On the other hand, NDCG evaluates ranking quality and relevance weighting, rewarding models that prioritize highly relevant items earlier in the list [27]–[29]. It is computed in Equation (8)

$$NDCG@k = \frac{\mathrm{DCG@k}}{IDCG@k} \tag{8}$$

where DCG@k is calculated using a logarithmic discount factor, ensuring that top-ranked items contribute more to the final score, as defined in Equation (9):

$$DCG@k = \sum_{j=1...k} \frac{2^{rel_j} - 1}{\log_2(j+1)} \quad (9)$$

rel_j represents the relevance score of the item at position *j* in the ranked list, where *rel_j* = 1 if the recommended item matches the ground-truth item, and 0 if otherwise.

These standard metrics are also applied within ART-Q to assess recommendation quality at the session, quartile, and antecedent levels, as detailed in the following explanations. The evaluation includes session-level HR which uses common HR@k metric as defined in (6). It measures whether the ground-truth next-item *y* is successfully identified within any $topk\text{-}Q_i$. It is defined in Equation (10):

$$HR@k = \begin{cases} 1, if\ \exists Q_i \text{ and } y \in Topk\text{-}Q_i \\ 0, \text{otherwise} \end{cases} \quad (10)$$

HR@k is counted as 1 if at least one quartile contains *y* within its top-k items. This metric evaluates the overall success of the model per session, ensuring that at least one ranked quartile includes the actual next-item prediction.

In addition, Quartile-Level HR Evaluation (Q-HR@k) is proposed to measure whether the ground-truth next-item appears within the top-k items in each quartile independently. This is formulated in Equation (11):

$$Q\text{-}HR@k = \begin{cases} 1, if\ y \in Topk\text{-}Q_i \\ 0, \text{otherwise} \end{cases} \quad (11)$$

Where Q−HR@k is computed separately for each quartile Q_i. Higher Q-HR@k scores indicate broader quartile relevance, ensuring that lower-ranked quartiles also contribute meaningfully to the recommendation process.

Antecedent-Level HR Evaluation (A-HR@k) is also considered. It measures how many antecedents within a session contribute to identifying the ground-truth next-item. This is defined in Equation (12):

$$A\text{-}HR@k(x_i) = \begin{cases} 1, & if\ \exists Q_i\ , y \in Topk\text{-}Q_i \text{ for } x_i \\ 0, \text{otherwise} \end{cases} \quad (12)$$

Where $x_i \in u$ represents an individual antecedent being observed. A-HR@k counts as 1 for each antecedent that successfully ranks y within its top-k predictions across quartiles. This metric evaluates the distribution of predictive contributions across antecedents, ensuring that recommendations are not overly dependent on a single antecedent.

It is worth noting that Q-HR and A-HR accommodate the model's ability to match relevant items across different quartiles through distinct antecedents. As a result, a single test session may have more than one Q-HR and A-HR. Consequently, the average Q-HR and A-HR across the entire test dataset can also exceed 1 if the total hits surpass the number of test sessions.

To assess ranking precision, ART-Q applies MRR@k on quartile-specific predictions. It applies MRR@k separately to each $topk\text{-}Q_i$ that contains the ground-truth next-item *y*, denoted as $MRR@k(topk\text{-}Q_i)$. The formula, adapted from Equation (7), is defined in Equation (13)

$$MRR@k(topk\text{-}Q_i) = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u(topk\text{-}Q_i)} \quad (13)$$

Here, $rank_u(topk\text{-}Q_i)$ is the position of the first correctly recommended next-item *y* within $Topk\text{-}Q_i$. The final session-level MRR@k score is determined by selecting the maximum $MRR@k(topk\text{-}Q_i)$ value obtained across four quartiles within that session. The overall session MRR@k is then computed as the average of all session-level MRR@k scores.

Similarly, $NDCG@k(topk\text{-}Q_i)$ measures ranking quality and relevance weighting within each quartile Q_i, ensuring that highly relevant items appearing earlier receive higher scores. This improves the ranking efficiency of recommendations. The formula, adapted from Equation (8), is defined in Equation (14)

$$NDCG@k(topk\text{-}Q_i) = \frac{DCG@k(topk\text{-}Q_i)}{IDCG@k(topk\text{-}Q_i)} \quad (14)$$

The principles of $DCG@k(topk\text{-}Q_i)$ and $IDCG@k(topk\text{-}Q_i)$ remain the same as the original version (Equation (9)) ; However, rel_j represents the relevance of item y, assigned a value of 1 if $y \in topk\text{-}Q_i$, and 0 otherwise. The final session-level $NDCG@k$ score is determined by selecting the maximum $NDCG@k(topk\text{-}Q_i)$ value obtained across four quartiles within that session. The overall session $NDCG@k$ is then computed as the average of all session-level scores.

3.5. Pseudocode of ART-Q

The pseudocode below outlines the main phases of the ART-Q framework, showing how quartile partitioning is applied in the recommendation process.

```
Pseudocode: ART-Q Recommendation Framework
Input:
  - T = Training dataset of user sessions
  - U = Test dataset of sessions
  - k = top-k recommendation required
  - p1, p2, p3, p4 = proportion of quartiles,
    with sum of pi = 1
Output:
  - REC # dictionary of all recommendation
  - HR@k, QHR@k, AHR@k, MRR@k,
    NDCG@k # Evaluation metrics

----- Modelling Phase -----
# incremental rule generation
1. Initialize:
   FI = {}  #frequent itemset dictionary
   R = {} # rule dictionary
2. For each session s in T do
     # find all frequent 1-itemsets FI
     For each single item x in s:
       if {x} not in FI: FI[{x}] = 1
       else: FI[{x}] += 1
     # find all frequent 2-itemsets FI
     For each pair {x, y} in s:
       if {x, y}in FI: FI[{x, y}] = 1
       else FI[{x, y}]  += 1

# Rule dictionary functions as a data model built
from training data

3. For 2-itemsets (x, y) in FI:
     support = FI[{x, y}]
     confidence = support / FI[{x}]
     weight w = confidence * FI[{y}]
     if x not in R : R[x] = {}
     R[x][y] = w
4. Return R and FI

----- Inference and Evaluation Phase -----
5. Initialize:
   HR=0, QHR=0, AHR=0, MRR=0,
   NDCG=0
   REC = {} # REC[sid][x][Qi] = Topk_Q, sid
   is the session ID

6. # Session loop
   For each session u in U do
   a. Split u into u_x = {x1,...,xi}, y = xi+1
   b. REC[sid] = {} # dict. for each sid
   c. # Reset per-session counters:
      REC[sid][x] = {} # dict. for each x
      quartile_hit = 0
      antecedent_hit = 0
      session_MRR = 0
      session_NDCG = 0
   d. # Antecedent loop
      For each antecedent x in u_x do
        Y = R[x]  # consequents of x
        Sort Y by w(x,y) in descending order
        Partition Y into quartiles Q = [Q1, Q2,
        Q3, Q4] with proportions p1..p4
   e.   # Quartile loop, for generating
        multiple top-k items across quartiles
        For each Qi in Q do
          Topk_Q = select top-k items in Qi
          REC[sid][x][Qi] = Topk_Q

          if y in Topk_Q:
            quartile_hit = 1
            antecedent_hit += 1

            # Quartile-specific MRR (Eq.
            13)
            rank = position(y in Topk_Q)
```

```
            MRR_Qi = 1 / rank
            # Quartile-specific NDCG (Eq. 14, 9)
            DCG = DCG(Topk_Q, y)
            IDCG = IDCG(y)
            NDCG_Qi = DCG / IDCG

            # Keep the maximum quartile score for this session
            session_MRR = max(session_MRR, MRR_Qi)
            session_NDCG = max(session_NDCG, NDCG_Qi)

        # End of Quartile loop
        QHR += quartile_hit
        AHR += antecedent_hit
        MRR += session_MRR
        NDCG += session_NDCG

      # End Antecedent loop
      if quartile_hit == 1: HR += 1
  #End of session loop
```

7. Normalize evaluation metrics:
   HR = HR / |U| # maximum = 1
   QHR = QHR / |U| # may exceed 1
   AHR = AHR / |U| # may exceed 1
   MRR = MRR / |U| # maximum = 1
   NDCG = NDCG / |U| # maximum = 1
8. Return REC and all evaluation metrics

3.6. Experimental Works Design

The objective of this experiment is to evaluate the performance of ART-Q on two datasets in comparison with baseline models. The experiment utilizes two datasets from distinct domains: (a) The YooChoose dataset (YOO), a public e-commerce transaction dataset containing user session data from the RecSys Challenge 2015; (b) The MPM dataset, which was collected from a digital news platform through cookie-based logging of user reading activities over a two-week period in August 2024. The cookies captured sequences of news article IDs accessed by readers, with session lengths varying from very short (one to three clicks) to very long, depending on user browser settings rather than program constraints. Although the dataset is private, it can be made available upon request to the authors for research purposes. It should be noted that the dataset may contain biases due to differences in cookie retention policies and the tendency of popular articles to appear more frequently in user sessions.

These datasets present unique challenges for session-based RS. YOO, as an e-commerce dataset, features goal-driven interactions where users make rapid purchasing decisions based on product descriptions and promotions [30]. In contrast, MPM represents digital news consumption, where user behavior is more exploratory, engaging with multiple articles in a single session based on evolving interests [1], [31]. Moreover, items in each domain exhibit distinct characteristics. In e-commerce, product availability fluctuates based on inventory levels and market trends, while in news platforms, article relevance evolves dynamically in response to editorial curation, breaking news, and audience engagement—factors that collectively shape their visibility and impact over time [32], [33].

For experimentation, each dataset is partitioned into training and testing sets. Table 1 summarizes the core characteristics of the datasets. For the YOO dataset, the train-test split used in this study follows the version provided on the official GitHub repository maintained by the original authors [34], to ensure consistency with prior comparative studies. In contrast, the MPM dataset was split into training and testing sets using an approximate ratio of 85:15. The minimum session length is two items; maximum session lengths vary per dataset. Two different treatments were applied to the MPM-train dataset: one using the original session data with a maximum length of 421, and the other by splitting it into sub sessions with a maximum length of 10. The training results were averaged from these two treatments.

Table 1 Summary of Datasets Characteristics

Dataset	Session Number	Item Number	Min-Max Session Length
YOO-train	1,410,210	28,450	2 - 200
YOO-test	1,606	2,603	2 - 84
MPM-train	17,244	31,226	2 – 421 (orig.), and 2 – 10
MPM-test	3,999	6,494	2 – 10

The baseline models for benchmarking include Simple AR, SKNN, VSKNN, GRU4Rec, NextItNet and SR-GNN. The Python implementations for these models are available on GitHub [35], maintained by the authors of a comparative study on session-based recommendation models [4], [11], [36]. For brevity, this benchmarking suite is referred to as SREC. Notably, NextItNet was also run using the original implementation provided by its authors on GitHub [34], as it employs a different data loading mechanism than the SREC suite. The proposed ART-Q framework is implemented in Python using Jupyter Notebook.

The configuration for each method follows the default settings provided by the respective authors, as summarized in Table 2. For neural models, 10 and 50 epochs were used to control the number of training iterations and the best-performing result was selected as the final outcome, while the learning rate (lr) determines the step size during optimization. Parameters such as layers, kernel size, and dilation reflect the architectural depth and the model's ability to capture sequential dependencies [37], [38]. In the traditional AR method, minsup = 1 means that any itemset appearing at least once is included in the recommendation process. No minconf threshold was applied, allowing even low-confidence rules to be considered [15]. For the KNN family—SKNN and VSKNN—the parameter k represents the number of neighboring sessions used, while n defines the number of candidate items considered for recommendation. VSKNN employs cosine similarity to measure session closeness and applies Inverse Document Frequency (IDF) weighting to reduce popularity bias and enhance recommendation relevance [4], [39]. All models are evaluated based on HR@20, MRR@20 and NDCG@20. In addition, Q-HR@20, and A-HR@20 are used exclusively to assess ART-Q.

Three complementary evaluation approaches were conducted in this study. First, six variations in quartile proportions were applied to ART-Q across the four segments: (1) V1: a uniform distribution with equal weight (0.25) assigned to all quartiles; (2) V2–V5: each variant emphasized the highest proportion for Q1–Q4 (0.7) with 0.1 assigned to the remaining quartiles; and (3) V6: a distribution derived from empirical observations, where a significant portion of target-item hits occurred in the third and fourth quartiles. This variation experiment was designed as a form of design-choice ablation, in which alternative quartile-weighting configurations were systematically tested to examine their impact on model performance. Similar ablation strategies have recently been employed in active learning and heuristic optimization to evaluate the effect of specific design decisions on overall effectiveness [40], [41].

Table 2 Models' Parameters applied to Datasets

Models	Parameters
GRU4Rec	Epoch = 10, 50, lr = 0.08, layers =100
NextItNet	Epoch = 10, 50, lr = 0.001, kernel size = 3, dilation = [1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4], dilation channel = 256
SR-GNN	Epoch = 10, 50, lr = 0.006, hidden_size = 100, out_size = 100
AR	*minsup* = 1, and no *minconf* applied
SKNN	k = 500, n = 5000
VSKNN	k = 500, n = 5000, sim=cosine, idf = 5
ART-Q-V1	p1, p2, p3, p4 = 0.25
ART-Q-V2 (Q1)	p1 = 0.7, p2 = 0.1, p3 = 0.1, p4 = 0.1
ART-Q-V3 (Q2)	p1 = 0.1, p2 = 0.7, p3 = 0.1, p4 = 0.1
ART-Q-V4 (Q3)	p1 = 0.1, p2 = 0.1, p3 = 0.7, p4 = 0.1
ART-Q-V5 (Q4)	p1 = 0.1, p2 = 0.1, p3 = 0.1, p4 = 0.7
ART-Q-V6	p1 = 0.01, p2 = 0.09, p3 = 0.5, p4 = 0.3

Second, the evaluation of baselines and the best-performing ART-Q variants on each dataset followed established practices in session-based recommendation research, such as GRU4Rec, SR-GNN, and NextItNet. In line with prior works, model performance was primarily evaluated using ranking-based metrics (HR, MRR, NDCG).

Finally, statistical significance testing was not conducted; instead, descriptive statistical analysis of the quartiles contributing most to metric improvements was provided, offering insight into ART-Q's behavior. Boxplots were used to depict the distribution of consequent sets and the distribution of hit rates across quartiles, thereby highlighting how different strata influence ranking quality. This complementary analysis reinforces the ablation findings by not only showing the effect of varying quartile weights but also uncovering the underlying distributional patterns that explain ART-Q's performance gains. Furthermore, this analysis revealed specific session cases where ART-Q failed to achieve a hit, providing insight into its limitations.

The SREC suite and ART-Q framework were executed on a laptop with an Intel Core i7 processor and a GeForce GTX 1650 GPU. In contrast, NextItNet was run on Google Colab using an A100 GPU in order to accelerate training on the YOO dataset, which proved slow on the local machine. Although the authors of NextItNet recommend splitting long sessions in the YOO dataset into sub-sessions [34], this strategy was ultimately not pursued due to excessive RAM usage, even on the Colab server.

III. RESULTS AND DISCUSSION

3.1. Results

The comparison of ablation results for ART-Q variants on the YOO and MPM datasets is presented in Figures 2 and 3. The HR, Q-HR, and A-HR metrics are combined in the same chart, while MRR and NDCG are shown separately.

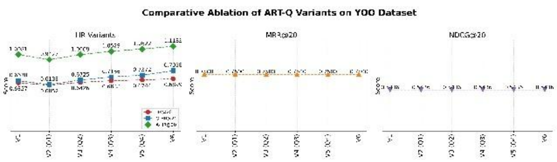

Figure 2 Comparative Ablation of ART-Q variants on YOO Dataset

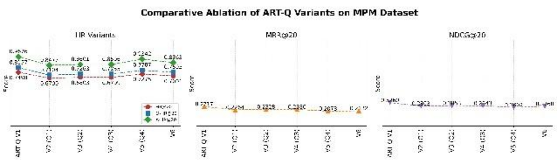

Figure 3 Comparative Ablation of ART-Q variants on MPM Dataset

On the YOO dataset, the ablation study of ART-Q variants demonstrates how different quartile weightings affect performance. With uniform proportions (V1: 0.25 each), ART-Q achieves a balanced baseline (HR@20 = 0.6357, Q-HR@20 = 0.6638, A-HR@20 = 1.0081). When the weighting is skewed toward the early quartiles (V2–V3, with Q1 or Q2 emphasized at 0.7), HR@20 declines slightly (0.6052–0.6426) and both Q-HR@20 and A-HR@20 fall below or close to the uniform baseline, indicating weaker generalization when recommendations rely heavily on top-ranked strata.

In contrast, shifting emphasis toward the later quartiles (V4–V5, with Q3 or Q4 emphasized at 0.7) consistently improves HR@20 (0.6638–0.6744), Q-HR@20 (0.7136–0.7372), and A-HR@20 (1.0529–1.0822). Further, the empirically derived distribution in V6 (0.01, 0.09, 0.5, 0.3) delivers the best overall performance, with HR@20 = 0.6899, Q-HR@20 = 0.7958, and A-HR@20 = 1.1183, confirming that stronger weighting of deeper quartiles enables ART-Q to retrieve relevant items that would otherwise be overlooked in traditional top-k approaches.

An A-HR value exceeding 1 indicates that multiple antecedents contribute to improving the hit rate across multiple consequent sets, surpassing the number of cases in the test sessions. This contribution highlights a key distinction from traditional AR, which relies solely on the last single antecedent to determine the next item. Through the use of multiple antecedents across quartiles, ART-Q is able to hit the ground-truth item more effectively.

On the MPM dataset, the ablation study shows that the uniform distribution (V1) delivers the highest overall performance, with the highest HR@20 (0.7498), MRR@20 (0.2717), and NDCG@20 (0.3260), while also achieving high Q-HR@20 (0.8157) and A-HR@20 (0.9676). Skewing the weighting toward Q1–Q3 (V2–V4) leads to consistent reductions across all metrics, indicating that overemphasis on early quartiles weakens the model's ability to capture the broader dynamics of news consumption.

Emphasizing Q4 (V5) improves HR@20 (0.7275) and yields relatively high Q-HR@20 (0.7787), but suffers in MRR@20 (0.2073) and NDCG@20 (0.2602), suggesting diminished ranking precision despite better retrieval. The empirically driven weighting (V6) produces moderate results (HR@20 = 0.7001, MRR@20 = 0.2132, NDCG@20 = 0.2760), outperforming skewed early-quartile variants but still lagging behind V1.

Overall, these findings confirm that balanced quartile weighting is effective in dynamic domains like news recommendation, where user interests are distributed more evenly across quartiles. Based on these results, V6 and V1 were selected for comparison with the respective baseline models on the YOO and MPM datasets. The comparative results are presented in Figures 4 and 5, respectively.

In brief, on the YOO dataset, non-neural models such as AR, SKNN, and VSKNN perform competitively, while neural models achieve slightly higher scores in some metrics but without consistent dominance. ART-Q-V6 attains the overall best performance across HR, MRR, and NDCG. On the MPM dataset, neural models show weak results, whereas non-neural methods perform better, with AR being the best baseline. ART-Q-V1, with uniform quartile distribution, achieves the best overall performance. The underlying reasons for these improvements are elaborated in the Discussion section.

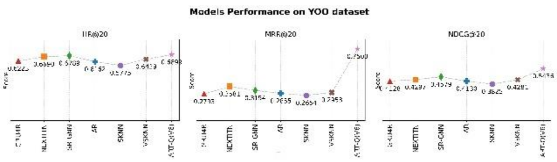

Figure 4 Model Performance on the YOO Dataset

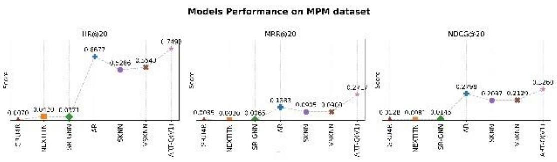

Figure 5 Model Performance on the MPM Dataset

3.2. Discussion

On the YOO dataset, non-neural models such as AR, SKNN, and VSKNN perform competitively, with VSKNN showing the highest results among them (HR@20: 0.6429, MRR@20: 0.2952, NDCG@20: 0.4281). Surprisingly, their performance on HR@20 matches or exceeds that of neural models like GRU4Rec, NextItNet, and SR-GNN which achieve HR@20 of 0.6223, 0.6690, and 0.6789 respectively.

Among the evaluated neural models, SR-GNN achieved the highest Hit Rate (HR@20 = 0.6789) and NDCG@20 (0.4579), demonstrating high capacity in capturing global session structure and long-range item transitions. NextItNet recorded the highest Mean Reciprocal Rank (MRR@20 = 0.3581), indicating superior precision in ranking ground-truth items at early positions—benefiting from its convolutional architecture for long-term dependencies. GRU4Rec, while earlier in design, remained competitive across all metrics (HR@20 = 0.6223, MRR@20 = 0.2793, NDCG@20 = 0.4128), validating its utility in modeling short-term sequential patterns. Overall, each neural model exhibited distinct strengths depending on the ranking objective, but none consistently dominated across all metrics.

In contrast, ART-Q-V6, with customized quartile proportions (p1 = 0.01, p2 = 0.09, p3 = 0.5, p4 = 0.3), consistently outperforms all baselines, achieving HR@20 = 0.6899, NDCG@20 = 0.5436, and a fixed MRR@20 = 0.7500. Relative to the traditional AR method, this corresponds to a +11.9% gain in HR, +31.6% in NDCG, and more than a two-fold improvement in MRR, highlighting ART-Q's effectiveness in ranking relevant items at earlier positions.

On the MPM dataset, neural models exhibited notably weak performance across all metrics. GRU4Rec achieved only HR@20 = 0.0070, MRR@20 = 0.0085, and NDCG@20 = 0.0128, indicating limited ability to capture relevant next-item signals in a non-sequential, rapidly evolving domain. NextItNet, although designed for long-range dependencies, recorded the highest hit rate (HR@20 = 0.0420) among the three, but suffered from the lowest MRR@20 = 0.0030, suggesting its recommendations were rarely positioned near the top. SR-GNN achieved slightly better NDCG@20 = 0.0145, reflecting modest improvements in ranking quality, yet remained suboptimal in both retrieval and ranking precision. These results highlight the challenge deep learning models face in dynamic domains like news consumption, where rigid sequential assumptions and static embeddings often fail to generalize to transient user interests and shifting content landscapes.

Non-neural models delivered markedly stronger performance than their neural counterparts on MPM dataset. The AR method achieved the highest results across all metrics—HR@20 = 0.6677, MRR@20 = 0.1383, and NDCG@20 = 0.2798—highlighting its robustness in capturing meaningful item transitions in dynamic content environments like digital news. While SKNN and VSKNN scored lower in retrieval (HR@20 = 0.5286 and 0.5543, respectively), they performed comparably in ranking quality (MRR@20 ~0.09; NDCG@20 ~0.21). Overall, these results confirm that classical rule-based approaches remain highly competitive for session-based news recommendation, especially when responsiveness and ranking consistency are prioritized over long-range sequence modeling.

Within ART-Q, among its variants, the uniform distribution of V1 produced the most balanced results (HR@20 = 0.7498, MRR@20 = 0.2717, NDCG@20 = 0.3260), suggesting that user interests in news consumption are best captured when candidate selection is spread evenly across quartiles rather than skewed toward specific strata. This highlights the importance of quartile design choices in adapting to domains characterized by dynamic and fast-changing content.

Q-HR and A-HR, introduced in this study, proved effective in capturing ART-Q's ability to leverage both quartile-level and antecedent-level contributions. Q-HR quantifies retrieval success across quartiles—highlighting how relevant items can be ranked within different quartile strata—while A-HR reflects the contribution of multiple antecedents within a session in improving ranking consistency.

The reason behind these achievements is revealed through analyses of consequent set distributions, rule weights, and hit-rate allocation across quartiles. As shown in Figure 6, consequent set sizes vary substantially, with YOO reaching up to 6,863 items (median: 92) and MPM up to 8,304 (median: 18), underscoring that a single antecedent may generate candidate pools spanning several orders of magnitude. Such variation makes naive top-k selection prone to overfitting toward popular consequents. Complementing this, Figure 7 shows that rule weights are heavily skewed in YOO (Q3 = 3,199) but more evenly distributed in MPM (Q3 = 492.6, median = 197.1), reflecting domain-specific dynamics: high popularity bias in e-commerce versus broader topical spread in news. These findings directly motivate ART-Q's quartile-based filtering, which distributes attention across multiple strata to reduce overconcentration on popular items.

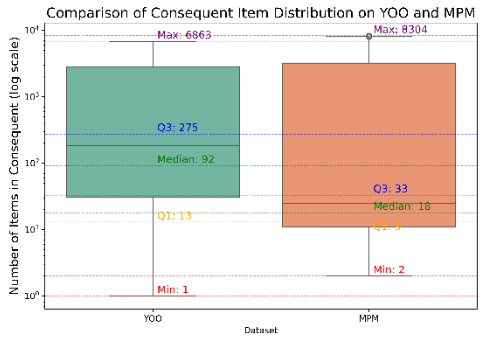

Figure 6 Comparison of Consequent Items Distribution on YOO and MPM

Figure 7 Comparison of Rule Weights Distribution on YOO (A) and MPM (B)

Figure 8 presents the heatmap of quartile contributions in retrieving the ground-truth item during inference. Each row corresponds to a test session, while columns represent the four quartiles. Blue cells denote successful hits and yellow cells indicate misses. The visualization highlights how the hit distribution is spread across quartiles, depending on the dataset and the ART-Q variant used. For the YOO dataset, results are shown for ART-Q-V6, while for MPM the visualization is based on ART-Q-V1, both being the best-performing configurations for their respective domains.

In YOO (V6), while a majority of hits occur in Q1 (56.8%), over 40% arise from Q2–Q4, showing that relevant items often lie beyond the highest-weighted rules. In MPM (V1), hits are balanced across all quartiles, with nearly half retrieved from Q3–Q4, highlighting the importance of deeper strata in dynamic domains. Interpreted through ablation, these patterns confirm that relying solely on Q1 reduces ART-Q to behavior resembling the traditional AR method—overlooking valuable signals from lower quartiles and leading to weaker retrieval and ranking quality.

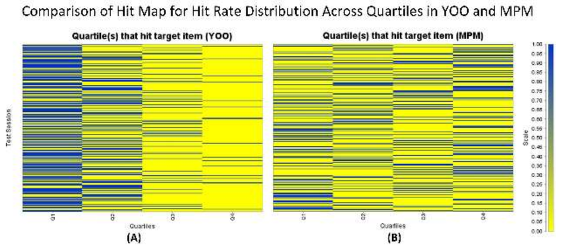

Figure 8 Comparison of Hit Map for Hit Rate Distribution Across Quartiles in YOO (A) and MPM (B)

On the other hand, ART-Q also encounters cases where test sessions fail to hit the ground-truth item. Although the train and test sets were prepared such that every test item appears in the training set to avoid cold-start scenarios, three conditions may still lead to misses. First, an antecedent x may lack a consequent set containing the ground-truth y because the pattern (x,y) never co-occurred in the training data. Second, the consequent set may contain uniform weights, with the ground-truth item falling outside the top-k and thus excluded from any quartile. Third, while the ground-truth y exists in the consequent set, its relatively small weight places it beyond the top-k. This condition was confirmed when k was increased (such as k = 50), and the ground-truth item was successfully covered.

Intuitively, the second and third cases could be mitigated through adaptive quartile proportions; however, the present version of ART-Q applies a static and uniform configuration across sessions, resulting in unavoidable misses. The first case corresponds to a classical cold-start scenario, which produces empty recommendations [14], [42]. In practice, this can be addressed by still offering the available top-k items within each quartile, since the system does not yet know the ground-truth next item in real-world applications. The top-k consequent items nevertheless retain contextual relevance with the antecedent, since their weights are derived from *confidence* and *support* measures. Even when the ground-truth item is absent, these candidates remain semantically plausible, reflecting observed co-occurrence patterns in the training data. This property suggests that ART-Q can still provide meaningful recommendations, even in cases where perfect accuracy is not achieved.

While this study primarily focused on ranking quality using standard metrics such as HR, MRR, and NDCG, aspects of recommendation diversity remain outside its current scope. Because the datasets used here consist only of anonymized item IDs, diversity could not be directly assessed. A more comprehensive evaluation would require incorporating item metadata—for example, product names and descriptions in the e-commerce domain or article titles and contents in the news domain. Such metadata would enable the measurement of content-level diversity, novelty and long-tail exposure, complementing the ranking-focused evaluation presented in this work.

From a resource utilization perspective, neural-based methods require significantly longer training times. GRU4Rec and SR-GNN demand several hours on a standard laptop GPU, while NextItNet requires substantial computational resources, taking approximately 3.5 hours on a Google Colab's A100 GPU. In contrast, non-neural approaches, including ART-Q, complete training and inference within less than two minutes on standard Laptop CPU, making them highly efficient and practical for large-scale deployments.

IV. CONCLUSION

This study introduced ART-Q, a quartile-based ranking framework designed to improve recommendation ranking quality in session-based recommendation systems. Across two domains—e-commerce (YOO) and digital news (MPM)—ART-Q consistently outperformed classical rule-based and KNN baselines, and achieved competitive results relative to neural models such as GRU4Rec, NextItNet, and SR-

GNN. By partitioning consequent sets into quartiles and applying multiple top-k recommendation across these strata, ART-Q reduced ranking ambiguity and increased the likelihood that relevant items appear earlier in the list. This strategy improved Hit Rate, Mean Reciprocal Rank, and NDCG, while maintaining a lightweight and computationally efficient design suitable for real-time applications. In addition, new evaluation metrics at the quartile and antecedent level were introduced, enabling finer-grained analysis aligned with the multiple top-k recommendation approach.

Nonetheless, several limitations must be acknowledged. First, the current implementation relies on static quartile partitioning, which may limit adaptability to evolving session dynamics. Second, the datasets used consist only of anonymized item IDs, preventing evaluation of content-level diversity or novelty beyond ranking-based metrics. Third, although training and testing splits ensured item overlap, ART-Q still encountered cases where ground-truth items were not retrieved, reflecting challenges related to cold-start and sparse co-occurrence patterns.

Future research can extend ART-Q in several directions. One avenue is the development of adaptive quartile weighting schemes that adjust dynamically to session characteristics or item popularity trends. Another is to incorporate content metadata (e.g., product descriptions, article titles) to evaluate diversity, novelty, and long-tail exposure more explicitly, since these aspects were beyond the scope of this study.

V. ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from the Directorate of Research, Technology, and Community Service (DRTPM), Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, under the 2024 Regular Fundamental Research Grant No. 109/E5/PG.02.00.PL/2024.

REFERENCES

[1] S. Raza and C. Ding, "News recommender system: a review of recent progress, challenges, and opportunities," *Artif. Intell. Rev.*, vol. 55, no. 1, 2022, doi: 10.1007/s10462-021-10043-x.

[2] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.

[3] X. Huang, Y. He, B. Yan, and W. Zeng, "Fusing frequent sub-sequences in the session-based recommender system[Formula presented]," *Expert Syst. Appl.*, vol. 206, 2022, doi: 10.1016/j.eswa.2022.117789.

[4] M. Ludewig, N. Mauro, S. Latifi, and D. Jannach, "Empirical analysis of session-based recommendation algorithms," *User Model. User-adapt. Interact.*, vol. 31, no. 1, 2021, doi: 10.1007/s11257-020-09277-1.

[5] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, and D. Lian, "A Survey on Session-based Recommender Systems," *ACM Comput. Surv.*, vol. 54, no. 7, 2022, doi: 10.1145/3465401.

[6] K. Wu and K. Chi, "Enhanced E-commerce Customer Engagement: A Comprehensive Three-Tiered Recommendation System," *J. Knowl. Learn. Sci. Technol. ISSN 2959-6386*, vol. 2, no. 3, 2024, doi: 10.60087/jklst.vol2.n2.p359.

[7] Z. Zhu *et al.*, "Understanding or Manipulation: Rethinking Online Performance Gains of Modern Recommender Systems," *ACM Trans. Inf. Syst.*, vol. 42, no. 4, 2024, doi: 10.1145/3637869.

[8] B. Hidasi and D. Tikk, "GRU4Rec: Session based recommendations with rnn," *Data Min. Knowl. Discov.*, vol. 30, no. 2, 2015.

[9] F. Yuan, A. Karatzoglou, I. Arapakis, J. M. Jose, and X. He, "A simple convolutional generative network for next item recommendation," in *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, 2019. doi: 10.1145/3289600.3290975.

[10] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural

networks," in *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, 2019. doi: 10.1609/aaai.v33i01.3301346.

[11] M. Ludewig, N. Mauro, S. Latifi, and D. Jannach, "Empirical analysis of session-based recommendation algorithms: A comparison of neural and non-neural approaches," *User Model. User-adapt. Interact.*, vol. 31, no. 1, 2021, doi: 10.1007/s11257-020-09277-1.

[12] Z. Wang, "Empowering Few-Shot Recommender Systems With Large Language Models-Enhanced Representations," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3368027.

[13] W. Li, Y. Cai, M. H. Hanafiah, and Z. Liao, "An Empirical Study on Personalized Product Recommendation Based on Cross-Border E-Commerce Customer Data Analysis," *J. Organ. End User Comput.*, vol. 36, no. 1, 2024, doi: 10.4018/JOEUC.335498.

[14] S. Z. U. Hassan, M. Rafi, and J. Frnda, "GCZRec: Generative Collaborative Zero-Shot Framework for Cold Start News Recommendation," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3359053.

[15] T. N. Dinh, B. D. Nguyen, X. D. Vu, and Q. D. Truong, "Improve Medicine Prescribing Performance Using Recommendation Systems," in *Communications in Computer and Information Science*, 2024. doi: 10.1007/978-981-99-7666-9_25.

[16] M. Ludewig and D. Jannach, "Evaluation of session-based recommendation algorithms," *User Model. User-adapt. Interact.*, vol. 28, no. 4–5, 2018, doi: 10.1007/s11257-018-9209-6.

[17] H. S. Sheu and S. Li, "Context-aware Graph Embedding for Session-based News Recommendation," in *RecSys 2020 - 14th ACM Conference on Recommender Systems*, Association for Computing Machinery, Inc, Sep. 2020, pp. 657–662. doi: 10.1145/3383313.3418477.

[18] R. Yeganegi, S. Haratizadeh, and M. Ebrahimi, "STAR: A session-based time-aware recommender system," *Neurocomputing*, vol. 573, Mar. 2024, doi: 10.1016/j.neucom.2023.127104.

[19] T. M. Akhriza and I. D. Mumpuni, "A Time-Window Approach to Recommending Emerging and On-the-rise Items," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, 2022.

[20] F. Antonello, P. Baraldi, E. Zio, and L. Serio, "A Novel Metric to Evaluate the Association Rules for Identification of Functional Dependencies in Complex Technical Infrastructures," *Environ. Syst. Decis.*, vol. 42, no. 3, 2022, doi: 10.1007/s10669-022-09857-z.

[21] J. H. Chowdhury, M. B. Hossain, M. S. Kaiser, and M. S. Arefin, "Performance Comparisons in Association Rule Mining Over Public Datasets," in *Lecture Notes on Data Engineering and Communications Technologies*, 2022. doi: 10.1007/978-981-16-6636-0_57.

[22] D. Liu, "Design of Data Mining System for Sports Training Biochemical Indicators Based on Artificial Intelligence and Association Rules," *Int. J. Data Min. Bioinform.*, vol. 1, no. 1, 2024, doi: 10.1504/ijdmb.2024.10060195.

[23] L. Wang, L. Gui, and P. Xu, "Incremental sequential patterns for multivariate temporal association rules mining," *Expert Syst. Appl.*, vol. 207, 2022, doi: 10.1016/j.eswa.2022.118020.

[24] J. Li and S. Bao, "Personalized Recommendation Method of E-Commerce Products Based on In-Depth User Interest Portraits," *Int. J. Inf. Technol. Web Eng.*, vol. 19, no. 1, 2024, doi: 10.4018/IJITWE.335123.

[25] Z. Chen, W. Gan, J. Wu, K. Hu, and H. Lin, "Data Scarcity in Recommendation Systems: A Survey," *ACM Trans. Recomm. Syst.*, vol. 3, no. 3, 2025, doi: 10.1145/3639063.

[26] N. Liu and J. Zhao, "Recommendation System Based on Deep Sentiment Analysis and Matrix Factorization,"

IEEE Access, vol. 11, pp. 16994–17001, 2023, doi: 10.1109/ACCESS.2023.3246060.
[27] S. Kanwal, S. Nawaz, M. K. Malik, and Z. Nawaz, "A review of text-based recommendation systems," *IEEE Access*, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 31638–31661, 2021. doi: 10.1109/ACCESS.2021.3059312.
[28] J. Polohakul, E. Chuangsuwanich, A. Suchato, and P. Punyabukkana, "Real Estate Recommendation Approach for Solving the Item Cold-Start Problem," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3077564.
[29] A. Gharahighehi and C. Vens, "Diversification in session-based news recommender systems," *Pers. Ubiquitous Comput.*, vol. 27, no. 1, 2023, doi: 10.1007/s00779-021-01606-4.
[30] G. D. S. P. Moreira, S. Rabhi, J. M. Lee, R. Ak, and E. Oldridge, "Transformers4Rec: Bridging the Gap between NLP and sequential/session-based recommendation," in *RecSys 2021 - 15th ACM Conference on Recommender Systems*, 2021. doi: 10.1145/3460231.3474255.
[31] P. Symeonidis, L. Kirjackaja, and M. Zanker, "Session-based news recommendations using SimRank on multi-modal graphs," *Expert Syst. Appl.*, vol. 180, 2021, doi: 10.1016/j.eswa.2021.115028.
[32] M. Karimi, B. Cule, and B. Goethals, "Leveraging Sequential Episode Mining for Session-Based News Recommendation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2023. doi: 10.1007/978-981-99-7254-8_46.
[33] A. Smets, J. Hendrickx, and P. Ballon, "We're in This Together: A Multi-Stakeholder Approach for News Recommenders," *Digit. Journal.*, vol. 10, no. 10, 2022, doi: 10.1080/21670811.2021.2024079.
[34] F. Yuan, "Generative model for sequential recommendation based on Convolution Neural Networks (CNN))." GitHub, 2024. [Online]. Available: https://github.com/fajieyuan/WSDM2019-nextitnet
[35] Malte and S. Latifi, "Python-based framework for building and evaluating session-based and session-aware recommender systems.," 2021. https://github.com/rn5l/session-rec
[36] S. Latifi, N. Mauro, and D. Jannach, "Session-aware recommendation: A surprising quest for the state-of-the-art," *Inf. Sci. (Ny).*, vol. 573, 2021, doi: 10.1016/j.ins.2021.05.048.
[37] Y. Ouyang, H. Long, R. Gao, and J. Liu, "Course Recommendation Model Based on Layer Dropout Graph Differential Contrastive Learning," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3352043.
[38] L. H. Baniata, S. Kang, M. A. Alsharaiah, and M. H. Baniata, "Advanced Deep Learning Model for Predicting the Academic Performances of Students in Educational Institutions," *Appl. Sci.*, vol. 14, no. 5, 2024, doi: 10.3390/app14051963.
[39] M. F. Dacrema, S. Boglio, P. Cremonesi, and D. Jannach, "A Troubling Analysis of Reproducibility and Progress in Recommender Systems Research," *ACM Trans. Inf. Syst.*, vol. 39, no. 2, Mar. 2021, doi: 10.1145/3434185.
[40] R. M. Fajri, A. Saxena, Y. Pei, and M. Pechenizkiy, "FAL-CUR: Fair Active Learning using Uncertainty and Representativeness on Fair Clustering[Formula presented]," *Expert Syst. Appl.*, vol. 242, 2024, doi: 10.1016/j.eswa.2023.122842.
[41] Y. Koguma, "Tabu Search-Based Heuristic Solver for General Integer Linear Programming Problems," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3361323.
[42] M. N. Noori, J. Ahamed, and M. Ahmed, "Matrix Factorization and Cosine Similarity Based Recommendation System For Cold Start Problem in E-Commerce Industries," *Int. J. Comput. Digit. Syst.*, vol. 15, no. 1, 2024, doi: 10.12785/ijcds/150156.

Appendix

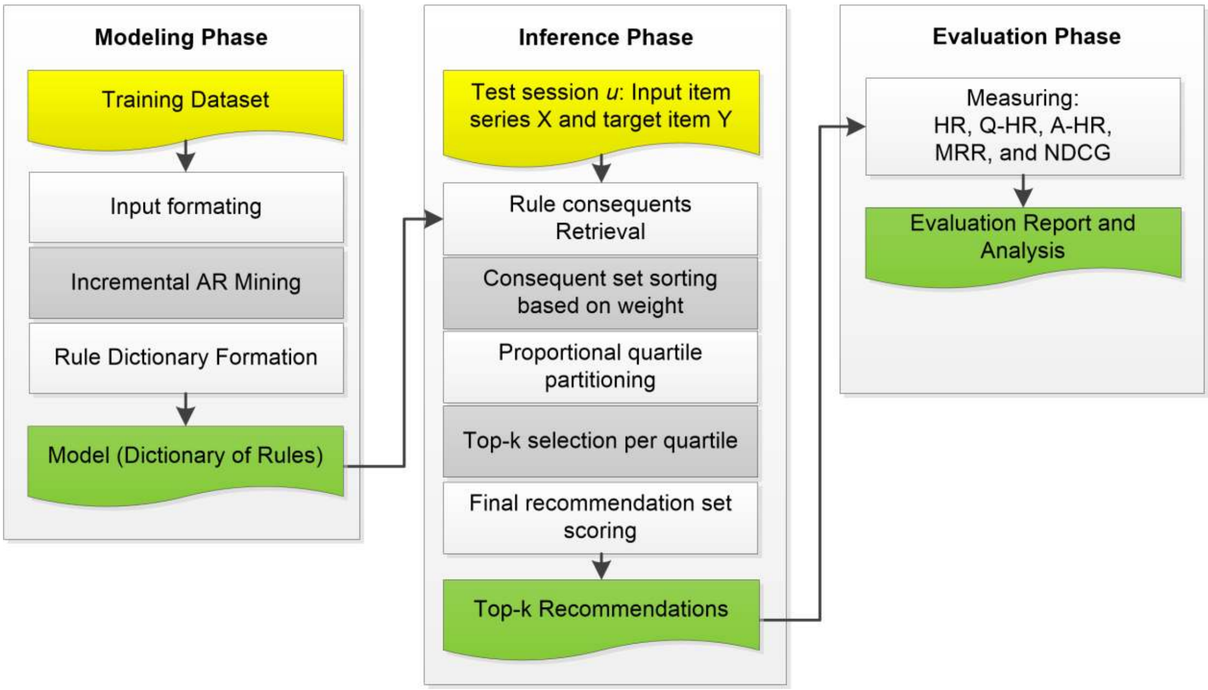

Figure 1. General Workflow of ART-Q method

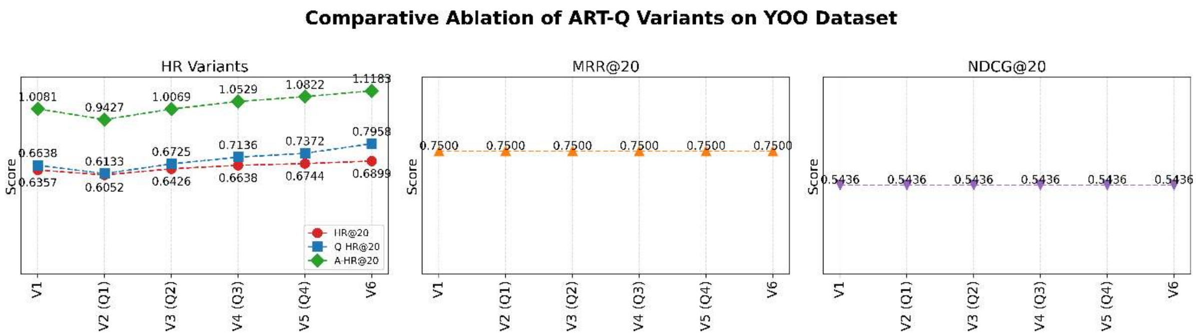

Figure 2 Comparative Ablation of ART-Q on YOO Dataset

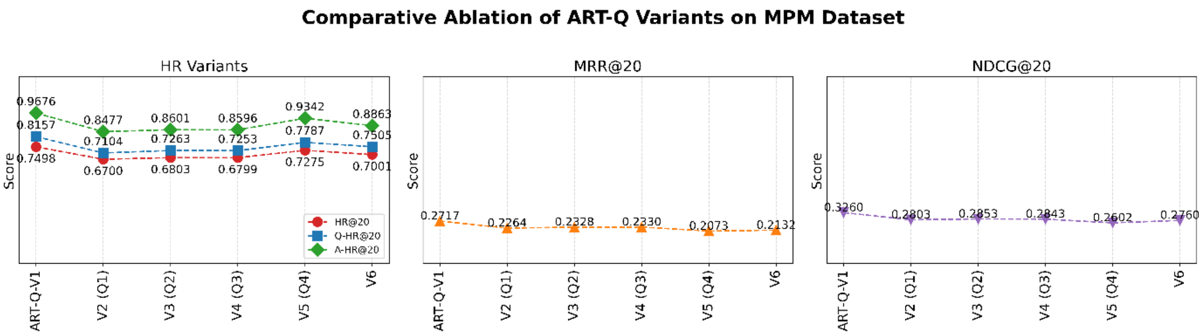

Figure 3 Comparative Ablation of ART-Q on MPM Dataset

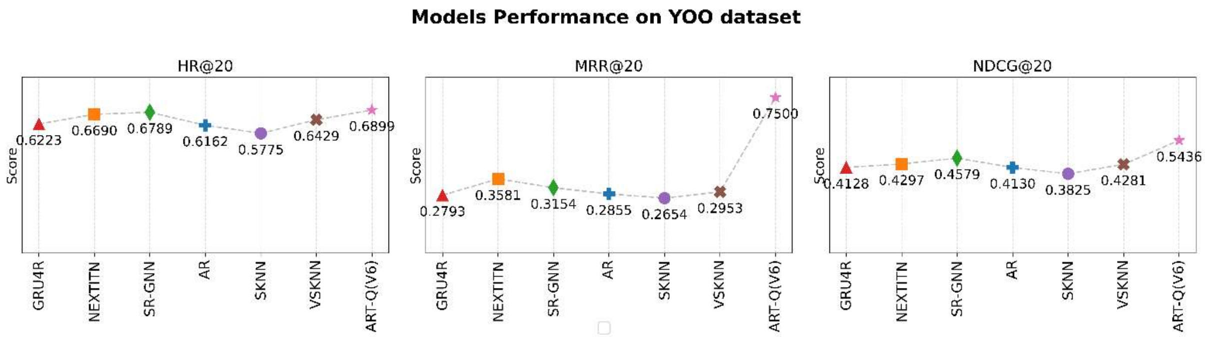

Figure 4 Model Performance on the YOO Dataset

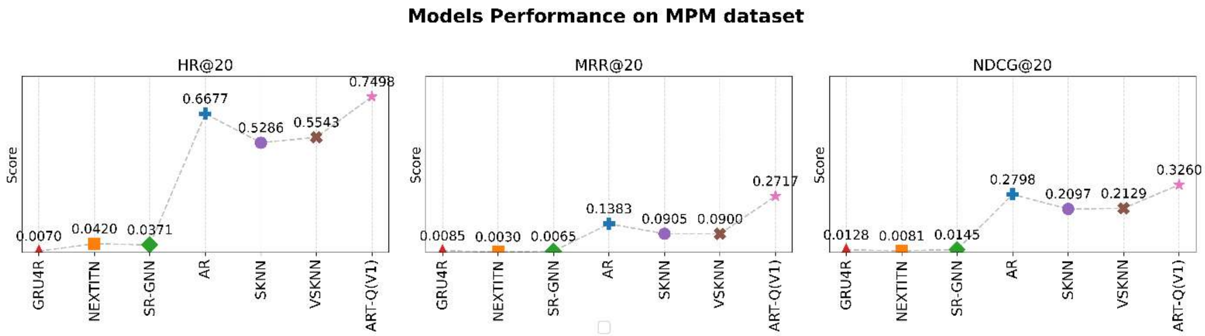

Figure 5 Model Performance on the MPM Dataset

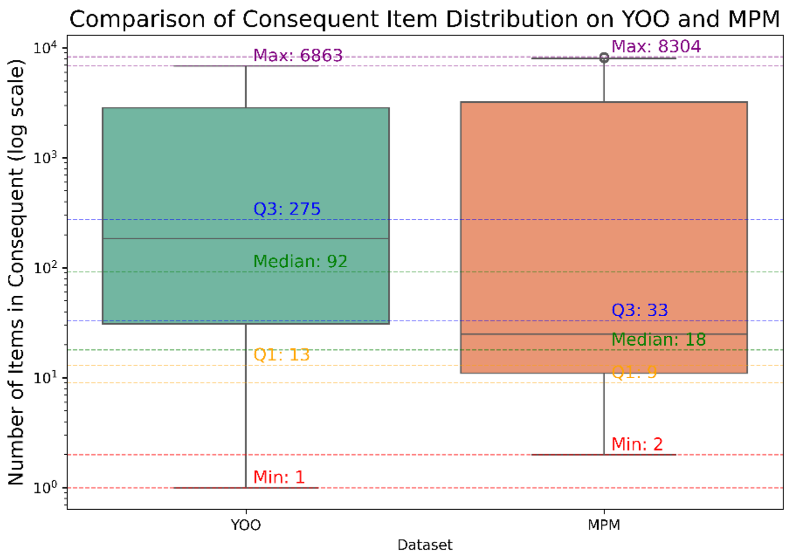

Figure 6 Comparison of Consequent Items Distribution in YOO and MPM

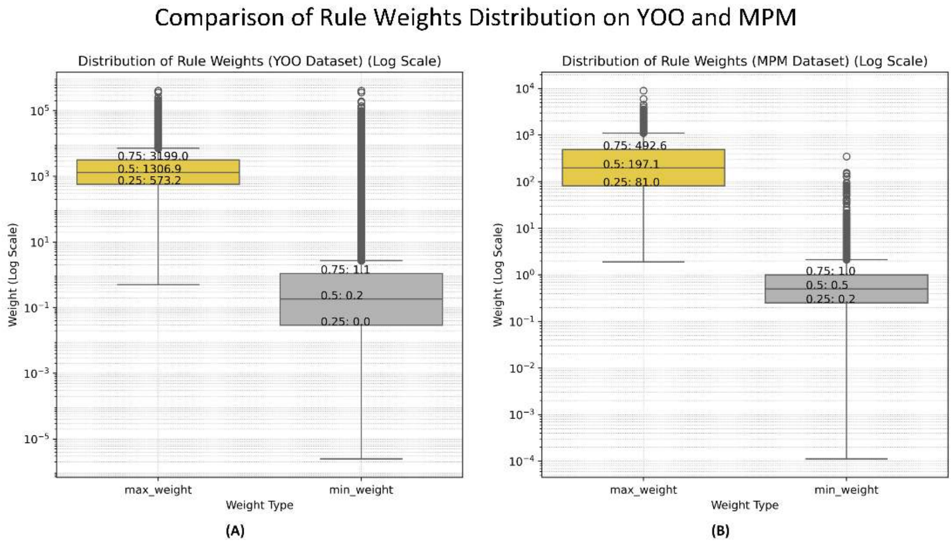

Figure 7 Distribution of Weights in YOO (A) and MPM (B)

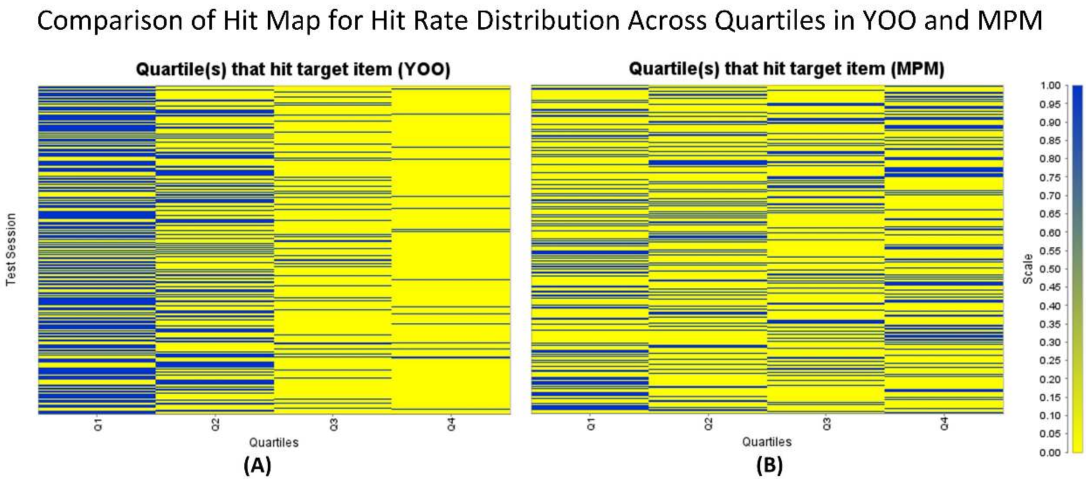

Figure 8 Heatmap of Quartiles that hit the target item in YOO (A) and MPM (B)

STIMATA
Menuju Kampus Riset 2020!

TB M Akhriza <akhriza@stimata.ac.id>

[CommIT] Editor Decision 2 - 13889

6 pesan

Eka Yanti Pangputri <epangputri@binus.edu> 29 Oktober 2025 pukul 13.41
Kepada: "akhriza@stimata.ac.id" <akhriza@stimata.ac.id>, "alqhoir@gmail.com" <alqhoir@gmail.com>, "weda@stimata.ac.id" <weda@stimata.ac.id>
Cc: CommIT Journal <commit@binus.edu>

Dear Mr/Mrs. Tubagus Mohammad Akhriza, Khoerul Anwar, and Weda Adistianaya Dewa,

We have reached a decision regarding your submission to CommIT (Communication and Information Technology) Journal, "QUARTILE-BASED FILTERING FOR ASSOCIATION RULE-BASED SESSION RECOMMENDATION".

Our decision is to: accept the article for October 2026 edition.

You can also add the **Acknowledgment** regarding your funding for the research in the article in **two weeks**. We may have deleted the information from your initial submission to conceal the authors' identity. The example can be seen in the template (attached). You can also inform us if you do not receive fund for this research.

Last, as your article is already accepted, please transfer the APC.

The article processing charge is Rp2.000.000. The payment information is as follow:

Bank Name: Bank Central Asia, cabang Bina Nusantara
Account Name: Univ Bina Nusantara
Account No: 5271706678

Thank you. Have a nice day.

Warmest Regards,

BINUS UNIVERSITY
BINUS RESEARCH EMPOWERMENT

Eka Yanti Pangputri

BINUS Journal Officer

Research Office | Research & Technology Transfer

BINUS UNIVERSITY | Anggrek Campus

t. (+62-21) 5345830 ext. 1708 | https://research.binus.ac.id/

Get in touch through our Social Media:

https://linktr.ee/RTTSocialMedia

2 lampiran

TEMPLATE Acknowledgement - CommIT - English.docx
27K

13889_REVISION.docx
1233K

Akhriza Tb M. <akhriza@stimata.ac.id> 30 Oktober 2025 pukul 10.59
Kepada: weda@stimata.ac.id, alqhoir@stimata.ac.id

[Kutipan teks disembunyikan]
--
Salam,
Dr. Tubagus M. Akhriza, S.Si., MMSI
Principal of STIMATA Campus, Malang

---

2 lampiran

TEMPLATE Acknowledgement - CommIT - English.docx
27K

13889_REVISION.docx
1233K

Akhriza Tb M. <akhriza@stimata.ac.id> 30 Oktober 2025 pukul 11.30
Kepada: Eka Yanti Pangputri <epangputri@binus.edu>
Cc: "alqhoir@gmail.com" <alqhoir@gmail.com>, "weda@stimata.ac.id" <weda@stimata.ac.id>, CommIT Journal <commit@binus.edu>

Dear Ms. Eka,

Thank you for the information.
I have uploaded the final version of the paper (including the acknowledgment) to OJS, and I have also sent the proof of payment to this email.

I have a few questions:
Since the paper is scheduled to be published in October 2026, is it possible for the *In Press* version to be released earlier?
Additionally, as the paper title has been changed from the original version, should I also update the metadata in OJS?

Thank you very much for your attention and assistance.

Best Regards,

Authors


ID #FT735887937

Transaksi berhasil diproses
30 October 2025 11:08:46 WIB

Rp2.000.000

Penerima	Univ Bina Nusantara
Bank Penerima	BCA
Nomor Rekening	5271706678
Catatan Transfer	
Pengirim	Tubagus Mohammad Akhriza
Transfer melalui	PT Fliptech Lentera

Kegiatan transfer uang di Flip terlisensi oleh Bank Indonesia.

[Kutipan teks disembunyikan]
[Kutipan teks disembunyikan]

Eka Yanti Pangputri <epangputri@binus.edu> 30 Oktober 2025 pukul 14.27

Kepada: "Akhriza Tb M." <akhriza@stimata.ac.id>
Cc: "alqhoir@gmail.com" <alqhoir@gmail.com>, "weda@stimata.ac.id" <weda@stimata.ac.id>, CommIT Journal <commit@binus.edu>

Dear Mr/Ms. Akhriza,

Thank you for the information.

1. For in-press, it will be hard to be done since we need to finish the May 2026 edition first.
2. We will update the metadata if it is already in editing process. For now, you cannot revise the title in OJS.

For your information, the next process may take long time since we are still prioritizing the journal that will be published before CommIT in 2026. Please make sure your email is active when we inform you the next process. **Moreover, please inform me your other emails (Yahoo or Gmail).**

We would also like to inform you that the accepted articles for CommIT Journal are required to use the service of professional proofreader and provide LaTeX as the final version source file. We may ask these files in the next year before the publication date.

If you have any questions, please let me know.

Thank you. Have a nice day

Warmest Regards,

Eka Yanti Pangputri

BINUS Journal Officer

RTT BINUS University

Get in touch through our Social Media:

https://linktr.ee/RTTSocialMedia

From: Akhriza Tb M. <akhriza@stimata.ac.id>
Sent: Thursday, October 30, 2025 11:31 AM
To: Eka Yanti Pangputri <epangputri@binus.edu>
Cc: alqhoir@gmail.com; weda@stimata.ac.id; CommIT Journal <commit@binus.edu>
Subject: Re: [CommIT] Editor Decision 2 - 13889

Dear Ms. Eka,

Thank you for the information.
I have uploaded the final version of the paper (including the acknowledgment) to OJS, and I have also sent the proof of payment to this email.

I have a few questions:
Since the paper is scheduled to be published in October 2026, is it possible for the *In Press* version to be released earlier?
Additionally, as the paper title has been changed from the original version, should I also update the metadata in OJS?

Thank you very much for your attention and assistance.

Best Regards,

Authors

ID #FT735887937

Transaksi berhasil diproses

30 October 2025 11:08:46 WIB

Rp2.000.000

Penerima	Univ Bina Nusantara
Bank Penerima	BCA
Nomor Rekening	5271706678
Catatan Transfer	

Pengirim	Tubagus Mohammad Akhriza
Transfer melalui	PT Fliptech Lentera

Kegiatan transfer uang di Flip terlisensi oleh Bank Indonesia.

[Kutipan teks disembunyikan]

Akhriza Tb M. <akhriza@stimata.ac.id> 30 Oktober 2025 pukul 15.13
Kepada: Eka Yanti Pangputri <epangputri@binus.edu>
Cc: "alqhoir@gmail.com" <alqhoir@gmail.com>, "weda@stimata.ac.id" <weda@stimata.ac.id>, CommIT Journal <commit@binus.edu>

Dear Ms. Eka,

Thank you for the clarification. You may also reach me at akhriza@gmail.com.

Best regards,
Akhriza

[Kutipan teks disembunyikan]

Eka Yanti Pangputri <epangputri@binus.edu> 30 Oktober 2025 pukul 18.51
Kepada: "Akhriza Tb M." <akhriza@stimata.ac.id>
Cc: "alqhoir@gmail.com" <alqhoir@gmail.com>, "weda@stimata.ac.id" <weda@stimata.ac.id>, CommIT Journal <commit@binus.edu>

Dear Mr/Ms. Akhriza,

Thank you for the information.

Warmest Regards,

[Kutipan teks disembunyikan]

CommIT (Communication and Information Technology) Journal

⟵ Back to Submissions

13889 / **Akhriza et al.** / QUARTILE-BASED FILTERING FOR ASSOCIATION RULE-BASED SESSION RECOMMENDATION

Library

Workflow | Publication

Submission | Review | Copyediting | Production

Round 1 | Round 2

Round 2 Status
Submission accepted.

Reviewer's Attachments — Search

No Files

Revisions — Search | Upload File

75377	13889_ACCEPTED_@CommIT UBinus.docx	October 30, 2025	Article Text

Review Discussions — Add discussion

Name	From	Last Reply	Replies	Closed
Next process	epangputri 2025-09-10 02:55 AM	akhriza 2025-10-02 08:43 AM	1	