Interdisciplinary Journal of Information, Knowledge, and Management

*An Official Publication of the Informing Science Institute InformingScience.org*

*IJIKM.org*

**Volume 21, 2026**

# Recurrent Naive Bayes for Multi-Criteria Recommender Systems: A Novel Approach for Partial Preference Imputation

| | | |
|---|---|---|
| Rita Rismala* | Telkom University, Bandung, Indonesia | ritaris@telkomuniversity.ac.id |
| Untari Novia Wisesty | Telkom University, Bandung, Indonesia | untarinw@telkomuniversity.ac.id |
| Febryanti Sthevanie | Telkom University, Bandung, Indonesia | sthevanie@telkomuniversity.ac.id |

* Corresponding author

## Abstract

| | |
|---|---|
| Aim/Purpose | This study aims to address the issue of partial preference, which causes incomplete criteria-level ratings and limits the accuracy of recommendations in multi-criteria recommender systems (MCRS). The primary focus of this study is to develop an imputation method that treats the imputation process as a sequential prediction task in which each missing criterion is predicted based not only on the known ratings but also on the previously imputed values. These preceding ratings serve as contextual input for the next prediction step, enabling the model to dynamically capture inter-criteria dependencies and thereby improve the overall imputation quality. |
| Background | A key challenge faced by MCRS is the issue of partial preference, which arises when users fail to provide ratings for all criteria. This results in incomplete criteria-level data, ultimately undermining the accuracy and effectiveness of MCRS in generating relevant recommendations. While most studies have primarily focused on improving the recommendation algorithms themselves, the pre-recommendation phase, which involves preparing high-quality and comprehensive input data, has often been overlooked. Yet, ensuring that multi-criteria data is as complete and accurate as possible is crucial for producing high-quality recommendations. This highlights the importance of a preprocessing step, particularly the imputation process to handle incomplete criteria ratings, applied prior to the recommendation generation process. |

Accepting Editor Dimitar Grozdanov Christozov │ Received: October 10, 2025 │ Revised: December 2, December 17, December 18, December 25, 2025; January 8, 2026 │ Accepted: January 9, 2026.
Cite as: Rismala, R., Wisesty, U. N., Sthevanie, F. (2026). Recurrent Naive Bayes for multi-criteria recommender systems: A novel approach for partial preference imputation. *Interdisciplinary Journal of Information, Knowledge, and Management, 21,* Article 3. https://doi.org/10.28945/5696

Rating-based imputation offers efficiency, while machine learning-based imputation offers accuracy. To leverage those advantages, this study employed rating-based imputation using a simple yet effective machine learning technique, Naive Bayes (NB). NB operates under the assumption of conditional independence among features, which may lead to suboptimal performance in contexts where semantic or perceptual correlations exist among criteria. In response to this limitation, this study introduces a sequential imputation approach, where each prediction of a missing criterion is informed by previously known or imputed ratings that are used as contextual input for the next prediction.

Methodology

This study introduces a new imputation method called Recurrent Naive Bayes (RNB), designed to estimate missing criteria ratings. Unlike traditional approaches, RNB models the imputation as a sequential prediction process, where each known or already imputed criterion $R_c$ serves as contextual information for predicting the subsequent missing criterion $R_{c+1}$. The core of RNB involves a recurrent process that includes prediction using NB, conditional imputation for the missing values, and updating the set of completed criteria. At each step, the model uses more features, starting with the overall rating and adding each newly predicted criterion one by one until all criteria are completed. In the recommendation process, RNB is used beforehand as part of the Multi-Criteria Collaborative Filtering (MCCF) pipeline. The performance of RNB is then evaluated by examining its impact on recommendation accuracy, specifically in terms of predicted criteria ratings, overall ratings, and recommended items, using three real-world datasets: TripAdvisor (TA), Yahoo! Movies (YM), and BeerAdvocate (BA).

Contribution

This study addresses the issue of partial preference in MCRS by highlighting the underexplored potential of NB for imputing missing criteria ratings. To this end, we propose RNB, a novel method that enhances traditional NB by adopting a sequential imputation strategy. Unlike conventional approaches that treat each criterion independently, RNB captures inter-criteria correlations, thereby improving imputation quality while preserving computational efficiency.

Findings

RNB consistently outperformed baseline methods significantly, including no imputation, mean imputation, and NB imputation, in enhancing MCRS performance across all datasets. This is demonstrated by its performance in predicting criteria ratings, overall ratings, and recommended items. The improvements were most notable in datasets with a larger number of criteria, stronger inter-criteria correlations, and larger size. Additionally, RNB enhanced recommendation accuracy for both moderately and highly recommended items.

Recommendations for Practitioners

By implementing MCRS, practitioners are encouraged to incorporate imputation techniques, such as RNB, during the preprocessing stage to improve data completeness and overall recommendation quality. RNB offers a practical balance between accuracy and efficiency, making it suitable for real-world applications where computational resources and data quality vary. It is especially recommended in scenarios involving sparse multi-criteria data with high inter-criteria correlation, such as in tourism, e-commerce, or entertainment domains.

Recommendations for Researchers

Adopting RNB can be beneficial for researchers as a practical and effective imputation method in MCRS, particularly when dealing with sparse and partially filled datasets. Its ability to capture inter-criteria dependencies through a sequential process makes it a valuable alternative to traditional imputation methods that assume feature independence. RNB's simplicity, efficiency, and ease of

| | |
|---|---|
| | integration into existing MCRS frameworks make it suitable for empirical studies that require scalable and interpretable preprocessing techniques. |
| Impact on Society | Recommender systems play a vital role in supporting decision-making across various domains such as tourism, entertainment, e-commerce, and education. By addressing the issue of partial preferences, RNB improves the completeness of user data, leading to more relevant and personalized recommendations. This impacts enhanced user satisfaction, better overall experiences, increased trust in digital platforms, and more informed consumer decisions. Moreover, the efficiency and simplicity of RNB support broader accessibility by enabling effective recommendations even in data-sparse environments. |
| Future Research | The performance of RNB depends on the assumption that inter-criteria dependencies can be captured in a sequential manner. However, the optimal ordering of criteria may differ between datasets, requiring dataset-specific analysis. Future research may explore dynamic or hybrid ordering strategies that simultaneously consider both sparsity and correlation to improve imputation quality. In addition, incorporating more advanced machine learning techniques could help model complex or nonlinear relationships among criteria. Furthermore, enhancing the robustness and adaptability of the imputation process. |
| Keywords | imputation, multi-criteria recommender system, partial preference, recurrent Naïve Bayes |

## INTRODUCTION

A recommender system (RS) has emerged as an essential technology for assisting users in making decisions based on their preferences. RS analyzes user behavior and historical data to provide personalized suggestions, improving the overall user experience (Al-Ghuribi & Mohd Noah, 2019; Isinkaye et al., 2015; Sahu et al., 2022). The application of RS has expanded rapidly across various domains, including e-commerce (Keikhosrokiani & Fye, 2024; Shankar et al., 2024), entertainment (Reddy et al., 2024; Shaw et al., 2024), tourism (Le et al., 2022; Shambour et al., 2024), education (Dehbozorgi et al., 2024; Du et al., 2025), and healthcare (Shambour et al., 2023; Yashudas et al., 2024). This widespread application reflects their growing importance in delivering context-aware and user-centric services in diverse environments.

Traditional RS typically relies on a single criterion, i.e., overall rating, to generate recommendations. This narrow focus limits its ability to fully understand the multifaceted nature of user preferences (Shambour, 2021). To address this limitation, Multi-Criteria Recommender System (MCRS) was introduced. MCRS extends the traditional RS by considering multiple criteria simultaneously (Adomavicius & Kwon, 2007). Therefore, MCRS can capture user preferences more comprehensively and provide more relevant recommendations (Nassar et al., 2020a, 2020b). One of the significant challenges encountered by MCRS is the issue of partial preference. This issue arises when users fail to provide ratings for all criteria (Hong & Jung, 2021b; F. Wang et al., 2013). Consequently, this can result in incomplete criteria ratings, ultimately undermining the accuracy of MCRS in generating the recommendations (Rismala et al., 2024b).

Most existing studies primarily focus on improving the recommendation algorithms themselves, often overlooking the crucial pre-recommendation phase, which involves preparing high-quality input data. In particular, the completion of missing criteria ratings is often neglected. However, having complete and accurate multi-criteria data is essential, as it can significantly improve the quality of recommendations (Rismala et al., 2024b). Therefore, a preprocessing method applied prior to the recommendation generation process is essential to prepare more complete and reliable input data, particularly by addressing missing criteria ratings.

In their study, Hong and Jung (2021b) identified four patterns of partial preference in MCRS: (1) users provide overall ratings without evaluating criteria; (2) users rate individual criteria but not the overall rating; (3) users leave ratings blank for their least preferred aspects; (4) users only rate some criteria while ignoring others. These patterns reflect the diverse ways in which users express their preferences, presenting significant challenges for imputing incomplete data accurately. Various imputation techniques have been proposed to address the issue of incomplete criterion ratings. Ranging from simply rating-based strategies to advanced review-based strategies that utilize machine learning techniques. The rating-based strategy is computationally efficient, but it has limitations in accuracy. For instance, Zhang et al. (2021) replaced missing values with the overall rating, Afzal et al. (2024) employed mean imputation, while Hong and Jung (2021b) applied a rule-based approach that considers the relationship between overall and criteria ratings. In contrast, the review-based strategies with machine learning techniques deliver enhanced accuracy; however, the usage of reviews typically requires more complex processing and higher computational resources. Included in these strategies, some studies incorporate several advanced techniques, such as unsupervised sentiment level analysis paired with a heuristic method (F. Wang et al., 2013), sentiment analysis utilizing the FlairElmo embedding (Hong & Jung, 2021a), Bidirectional Encoder Representations from Transformers (BERT)-based rating prediction (Zhuang & Kim, 2021), and the latest method, which features a multi-attribute BERT integrated with rule-based adjustments (Rismala et al., 2023, 2024b).

From those studies, we identified that rating-based strategies offer efficiency, while machine learning approaches contribute to predictive accuracy. Motivated by this, our study bridges both advantages by employing a rating-based strategy within a machine-learning framework. In machine learning techniques, Naïve Bayes (NB) is one of the popular methods for data imputation due to its simplicity, computational efficiency, and ability to handle high-dimensional data (Alam et al., 2023; Khotimah et al., 2019). NB, a probabilistic classifier based on Bayes' Theorem, assumes conditional independence among features. In the context of multi-criteria imputation, each criterion can be treated as a feature, allowing NB to estimate the probability of a missing criterion value based on known ratings. However, to the best of our knowledge, the application of NB in the context of multi-criteria rating imputation remains underexplored, particularly when considering inter-criteria dependencies.

Building on this gap, a key limitation of traditional NB lies in its assumption of feature independence, despite real-world user preferences often exhibiting semantic or perceptual correlations. For example, cleanliness may influence perceptions of service or room quality in hotel recommendations. Ignoring these correlations may lead to suboptimal imputations. Although Bayesian Networks (BNs) can explicitly model such dependencies, they require complex structure learning and large datasets, making them less scalable for real-world recommender systems (Kitson et al., 2023). This gap highlights the need for a method that preserves NB's efficiency while addressing inter-criteria relationships. To overcome this limitation, we introduce a novel approach called Recurrent Naive Bayes (RNB). RNB treats the imputation process as a sequential prediction task, where each known or previously imputed criterion is used as contextual input for the next prediction. This structure enables the model to capture inter-criteria relationships while maintaining computational efficiency, thereby improving MCRS performance.

Given these motivations, the main objective of this study is to investigate the impacts of RNB-based imputation on MCRS performance. Therefore, the research question guiding this study is: How does RNB-based imputation improve MCRS performance? To address this question, the remainder of this paper is organized as follows. The next section presents the proposed RNB method for multi-criteria rating imputation. The system implementation is then described, including the hardware and software environment as well as the training and testing procedures. The evaluation method is then outlined, covering the dataset, evaluation metrics, and experimental setup. The experimental results are then reported for both the criteria rating prediction and the recommended item prediction, followed by the key findings in relation to the research question. The findings are discussed further by providing comparative analyses and implications. The final section concludes the paper and highlights promising directions for future research.

# METHODOLOGY

Let U = {$u_1$, $u_2$, …, $u_m$} denote the set of users and I = {$i_1$, $i_2$, …, $i_n$} the set of items. In MCRS, each user u∈U may assign ratings to an item i∈I based on multiple criteria C = {$c_1$, $c_2$, …, $c_k$}, in addition to an overall rating $c_0$∈C. Our goal is to predict missing criterion-level ratings and restore the complete utility profile $R_{(u,i)}$ for each user-item pair.

## *RECURRENT NAIVE BAYES*

Partial preference problems, where only a subset of criteria is rated, are a major challenge in MCRS. To address this problem, our study proposed a novel imputation approach, namely Recurrent Naive Bayes (RNB), for estimating the missing criteria rating. RNB treats the imputation process as a sequential prediction task, where each known or previously imputed criterion $R_c$ is used as contextual input for the next prediction $R_{c+1}$. The framework of RNB is illustrated in Figure 1.

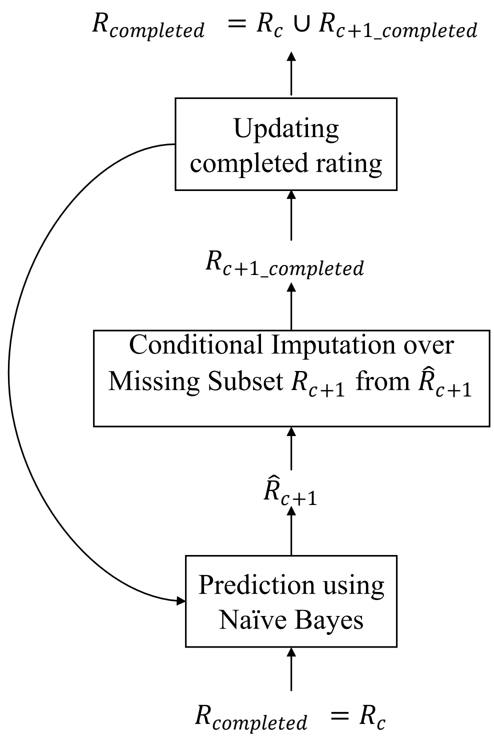



**Figure 1. RNB framework**

RNB performs iterative imputation by modeling each missing criterion rating conditionally using an NB classifier. At each step, the model leverages an expanding set of features, starting with the overall rating and sequentially augmented with each newly imputed criterion. The overall rating serves as the initial feature because it is typically the most complete criterion and provides a global summary of user preferences. This global signal guides the first-stage RNB predictions before other criteria are available. This approach assumes that the overall rating is broadly related to all sub-criteria, making it an appropriate initial dependency in the sequential imputation process. Furthermore, RNB assumes that the rating distributions can be modeled probabilistically using NB and that the sequential ordering of criteria influences the imputation flow. Through this recurrent mechanism, RNB enables the capture of dependency between criteria and uses predictions from earlier stages as context for the next stages.

### Prediction using Naive Bayes

Naive Bayes (NB) classifier is trained on the currently available ratings $R_c$ and previously imputed criteria, which initially include only the overall rating. The model is then used to predict the missing values for the next target criterion, denoted as $\hat{R}_{c+1}$.

NB serves as a simple yet powerful probabilistic classifier that can be effectively adapted for imputation tasks in MCRS. Its primary advantage lies in its ability to model the conditional distribution of each missing criterion rating given other observed features, under the assumption of conditional independence. This assumption, though simplifying, often yields surprisingly effective results in high-sparsity environments. Formally, given a missing criterion $R_c$ and a set of observed variables $X = \{x_1, x_2, \dots, x_n\}$, NB model estimates the probability distribution, as in Eq. (1). Then, the rating $r$ with the highest posterior probability is selected as the imputed value, as in Eq. (2).

$$P(R_c = r|X) \propto P(R_c = r) \prod_{j=1}^{n} P(x_j|R_c = r) \tag{1}$$

$$\hat{r} = arg \max_{r} P(R_c = r|X) \tag{2}$$

In our framework, the set $X$ expands with each iteration to include previously imputed criteria. For instance, the first missing criterion may be predicted using only the overall rating, while subsequent criteria can incorporate both the overall rating and earlier predictions, thereby enriching the conditional context. This recurrent use of NB enables the model to leverage the interdependencies among criteria in a step-wise fashion, compensating for missing data and reducing overall sparsity in the rating matrix.

### Conditional imputation

The predicted values $\hat{R}_{c+1}$ are used to impute the missing entries in the corresponding criterion $R_{c+1}$. This step, referred to as "Conditional Imputation over Missing Subset", ensures that only the missing values in $R_{c+1}$ are updated, while known values remain unchanged. Formally, for all $i \in I_{miss}^{c+1}$, we update $R_{c+1} \leftarrow \hat{R}_{c+1}$.

### Updating completed ratings

The imputed criterion $R_{c+1}$ is then appended to the feature set and treated as a completed attribute. This updated set of features is used in the subsequent iteration to predict the next missing criterion.

The process is recurrent in nature, as each newly imputed criterion is reused as input in the next predictions, thus progressively enriching the feature set. This allows the model to refine its understanding of user preferences over multiple dimensions and improves the quality of imputation as more context is incorporated. The overall pseudocode of RNB is presented in Algorithm 1, detailing the sequential steps of training, prediction, and conditional imputation across all rating criteria.

## *MULTI-CRITERIA COLLABORATIVE FILTERING*

Multi-Criteria Collaborative Filtering (MCCF) is the most extensively developed technique within MCRS. The main steps in MCCF typically involve predicting criteria ratings and overall ratings for items that have not yet been consumed (unknown items). The recommendation framework in MCCF, as adapted from Adomavicius and Kwon (2007), is illustrated in Figure 2. We propose a preprocessing method that is applied before executing the MCCF framework to improve data completeness and, consequently, recommendation accuracy, referred to as RNB.

Following the MCCF framework in Figure 2, the overall process of our proposed MCCF that incorporates the preprocessing step using RNB is depicted in Figure 3.

**Algorithm 1. RNB pseudocode**

```
Input
  R           // Dataset with missing multi-criteria ratings
  C           // List of criteria to impute [c1, c2, ..., ck]; sorted by lowest sparsity
  H           // Initial feature set ← {Overall}
  NB          // Naive Bayes classifier

Output
  R_completed        // Dataset with imputed ratings

procedure RNB (D, C, H, NB)
  R_completed ← R

  for each criterion c in C:
    // Step 1: Partition data
    R_train ← rows in R_completed where rating c is observed
    R_test ← rows in R_completed where rating c is missing

    // Step 2: Define target and feature set
    y_train ← column c in R_train
    X_train ← columns in H from R_train
    X_test  ← columns in H from R_test

    // Step 3: Train Naive Bayes model
    NB_model ← NB.train(X_train, y_train)

    // Step 4: Predict missing values
    y_pred ← NB_model.predict(X_test)

    // Step 5: Update completed dataset
    for all missing values in R_completed column c:
      replace with values from y_pred

    // Step 6: Update feature set
    H ← H ∪ {c}
  end for

  return R_completed
end procedure
```

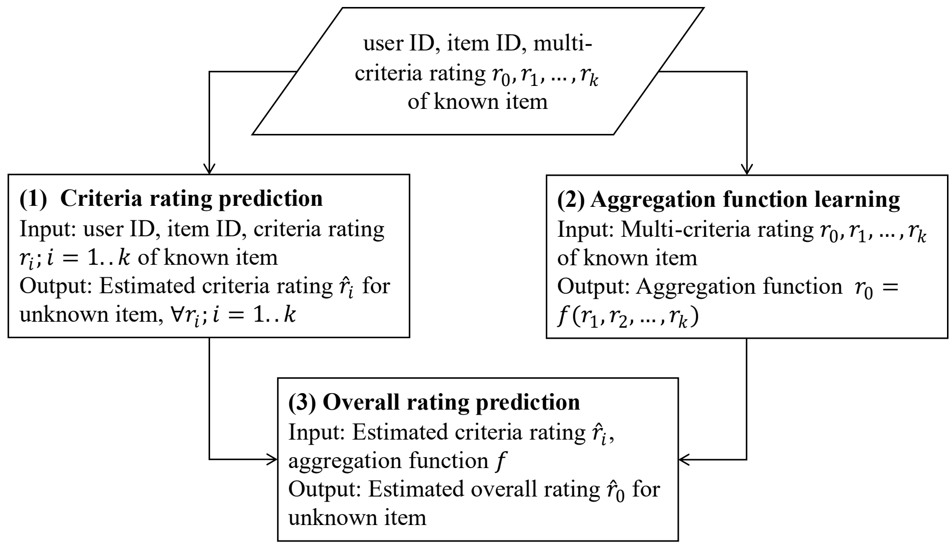



**Figure 2. MCCF framework (Adomavicius & Kwon, 2007; Rismala et al., 2024a)**

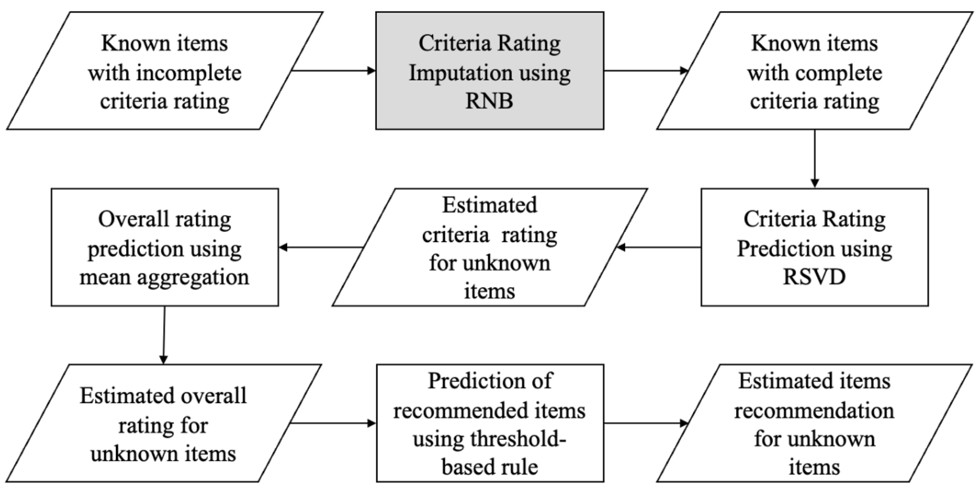
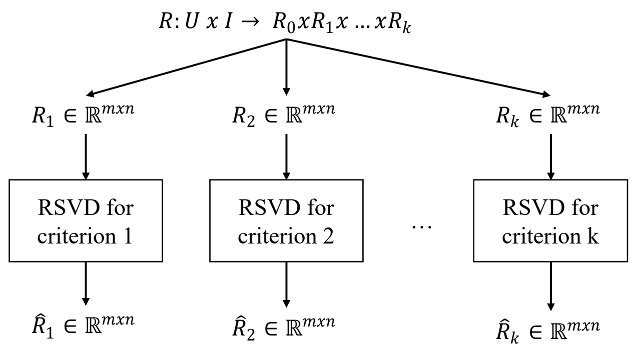



**Figure 3. Proposed MCCF with RNB**

In this context, we employed RNB for criteria rating imputation in the preprocessing stage. Then, to evaluate the effectiveness of the proposed RNB method, we integrated it into the standard MCCF framework. For the criteria rating prediction step, we adopted Regularized Singular Value Decomposition (RSVD) as presented in Figure 4 (Rismala et al., 2024a), while the overall rating was predicted using mean aggregation over the criteria ratings as in Eq. (3) (Shambour, 2021). The overall rating is then used to determine the recommended items by applying a threshold-based decision rule. Specifically, an item is considered recommended if its estimated overall rating $\hat{r}_0$ exceeds a certain threshold value. In this study, we define two threshold values, i.e., 3 and 4, which correspond to moderately and highly recommended items, respectively (Bag et al., 2019; Rismala et al., 2024a). These thresholds are selected because they align with common practices in rating-based RS, where ratings of 3 and above typically indicate acceptable user preference, and ratings of 4 and above reflect strong positive preference.

$$\hat{r}_0 = \frac{r_1 + r_2 + \cdots + r_k}{k} \tag{3}$$




*m: number of users; n: number of items*

**Figure 4. RSVD framework for criteria rating prediction**

## SYSTEM IMPLEMENTATION

### *HARDWARE ENVIRONMENT*

All experiments were conducted on the Google Colab cloud-based platform (free tier) without GPU acceleration. The environment was chosen based on the probabilistic characteristics of the proposed model. Therefore, it does not require large-scale parallel computing. The hardware specifications used are as follows:

1. CPU: Single-core virtual machine (Intel Xeon, 2.20 GHz, allocated by Colab).
2. RAM: ±12 GB (dynamically allocated on the Colab free tier).
3. Storage: ±70 GB of temporary storage space.

This configuration confirms that the proposed method does not require high computing resources. It can also be easily replicated by other researchers using widely available free services.

### *SOFTWARE ENVIRONMENT*

The whole implementation was conducted using the Python 3.10 programming language on Google Colab. The main libraries used were:

1. scikit-surprise (Surprise) for implementing recommender system algorithms, such as RSVD.
2. NumPy and Pandas for numerical operations and data manipulation.
3. scikit-learn for implementing machine learning methods, such as Naive Bayes, and evaluation functions, such as Mean Absolute Error (MAE) and F1-score.
4. Matplotlib and Seaborn for visualizing experimental results.

Library versions were locked during the experiment to ensure reproducibility: Surprise 1.1.4 and NumPy 1.26.4.

### *TRAINING AND EVALUATION PROCEDURE*

To evaluate model performance, each dataset was divided into 70% training data and 30% testing data. The splitting process used five random states to minimize bias in data selection and achieve consistent results. The training data was used to calculate prior probabilities and other probabilistic parameters required by the RNB model. While the testing data was used to measure the model's performance in three main aspects:

1. *Criteria rating prediction* – assesses the model's accuracy in estimating the rating for each criterion.
2. *Overall rating prediction* – tests the model's performance in estimating the overall rating based on the criteria ratings.

3. *Recommended item prediction* – assesses the model's performance in classifying recommended and non-recommended items.

The evaluation results are reported as the average performance of five observations with different random states. Furthermore, a paired t-test was conducted to assess the significance of RNB's performance compared to the baseline method at a 95% confidence level.

# EVALUATION METHOD

## *DATASET*

Experiments were conducted using real-world datasets from three different domains: TripAdvisor (TA) (H. Wang et al., 2010, 2011), Yahoo! Movies (YM) (Jannach et al., 2012), and BeerAdvocate (BA) (McAuley & Leskovec, 2013; McAuley et al., 2012). These datasets were selected based on three primary criteria: (1) their relevance to the multi-criteria rating imputation problem, (2) the availability of standardized evaluation settings, and (3) their accessibility as open-source resources. Beyond these factors, the datasets were also chosen to represent distinct domains with different rating behaviors, allowing the evaluation of RNB across diverse preference structures and sparsity patterns. Using datasets with heterogeneous characteristics helps reduce domain-specific bias and strengthens the generalizability of the findings, as consistent performance across different domains indicates robustness in varying multi-criteria contexts.

All datasets share a common structure comprising user ID, item ID, overall rating, and multi-criteria ratings. The TA dataset, representing the hotel domain, includes six criteria: service (c1), cleanliness (c2), value (c3), sleep quality (c4), rooms (c5), and location (c6). The YM dataset pertains to the movie domain and contains four criteria: story (c1), acting (c2), direction (c3), and visuals (c4). The BA dataset, from the beer domain, also includes four criteria: appearance (c1), aroma (c2), palate (c3), and taste (c4).

To reduce sparsity, a minimum thresholding was applied on user-item interactions. Both the TA and YM datasets include users and items with at least five interactions, while BA requires a minimum of ten interactions. All datasets use a rating scale from 1 to 5, with 5 representing the highest preference level. Notably, the original YM dataset used a 13-level alphanumeric grading scale (A+ to F). This scale was first converted to a numerical 13-point scale (1 to 13), where 13 corresponds to A+. To maintain consistency across datasets and facilitate comparative analysis, the ratings were subsequently normalized to match the 1–5 rating interval used in TA and BA.

A summary of the datasets, including key statistics such as the number of users, items, and rating sparsity, is presented in Table 1. In addition, the correlation between criteria for each dataset is visualized in Figure 5, providing insights into the inter-dependencies among the multi-criteria ratings.

**Table 1. Datasets summary**

| Properties | TA | YM | BA |
|---|---|---|---|
| #Users | 41,141 | 6,053 | 3,204 |
| #Items | 4,698 | 907 | 2,328 |
| #Overall ratings | 299,891 | 61,858 | 157,683 |
| Overall sparsity | 99.8448% | 98.8733% | 97.8860% |
| C1 sparsity | 99.8571% | 98.9576% | 98.4944% |
| C2 sparsity | 99.8568% | 99.4917% | 97.9604% |
| C3 sparsity | 99.8565% | 99.1077% | 98.5446% |
| C4 sparsity | 99.9197% | 99.3110% | 99.0585% |
| C5 sparsity | 99.8700% | - | - |
| C6 sparsity | 99.8765% | - | - |

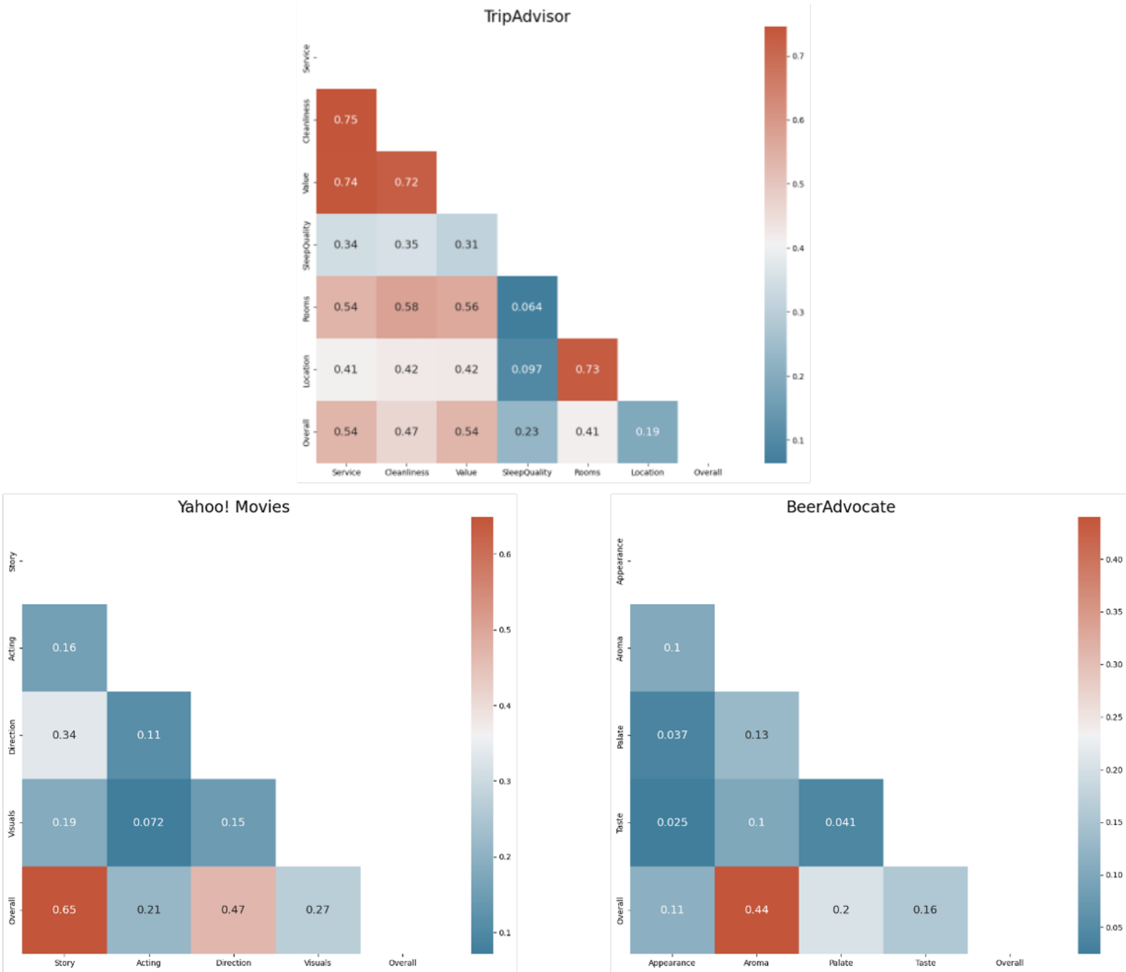



**Figure 5. Criteria correlation**

## *Performance Metrics*

The effectiveness of RNB is evaluated based on its impact on recommendation accuracy, including the predicted criteria rating, overall rating, and recommended items. The accuracy of criteria and overall rating predictions is measured using MAE, while the recommended item prediction is assessed using the Macro F1-score, as defined in Eq. (4) and (5), respectively. MAE measures the average absolute difference between the actual rating $y_i$ and the predicted rating $\hat{y}_i$. Meanwhile, the Macro F1-score is employed to assess the relevance of item recommendations by evaluating the classification performance across both the recommended and non-recommended classes equally, regardless of class imbalance.

$$MAE = \frac{1}{n}\sum_{1}^{n}|y_i - \hat{y}_i| \tag{4}$$

$$Macro\ F1 = 1/n\sum_{x} F1_x \tag{5}$$

In the Macro F1-score, *n* denotes the number of classes, and $F1_x$ represents the F1-score of class *x* (Opitz & Burst, 2021). In this study, the number of classes is two, corresponding to the recommended and not recommended categories. The F1-score for each class is calculated using Eq. (8), based on the confusion matrix shown in Table 2. This matrix includes True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

**Table 2. Confusion matrix**

| | | Predicted value | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Actual value** | **Positive** | TP | FN |
| | **Negative** | FP | TN |

$$precision = TP/(TP + FP) \tag{6}$$

$$recall = TP/(TP + FN) \tag{7}$$

$$F1 = (2 \times precision \times recall)/(precision + recall) \tag{8}$$

### *EXPERIMENTAL SETUP*

**RSVD parameter setting.** The feature dimension and the learning rate are two key hyperparameters in the RSVD algorithm. The RSVD model was evaluated using default parameters, with the feature dimensions set to 50 and the learning rate set to 0.1.

**RNB parameter setting.** In this study, we adopt an ordering strategy based on sparsity, where criteria with the lowest missing rate (i.e., most complete data) are predicted first. This decision is grounded in the recurrent nature of the RNB model, in which the output of each imputed criterion becomes part of the input for subsequent predictions. By prioritizing criteria with more complete data, the model benefits from more reliable statistical estimates in the early stages of the imputation process. This helps minimize error propagation across subsequent predictions, enhances the stability of conditional probability estimations, and reduces uncertainty in early-stage imputations that serve as the foundation for later steps.

For the TA dataset, the criteria were ordered as value, cleanliness, service, rooms, location, and sleep quality. For the YM dataset, the ordering was story, direction, visuals, and acting. Meanwhile, for the BA dataset, the criteria were arranged as aroma, appearance, palate, and taste.

**Baseline methods.** To evaluate the effectiveness of RNB, we compared its recommendation performance against three baseline approaches: (1) no imputation (NI), (2) mean imputation (Afzal et al., 2024), and (3) Naïve Bayes (NB)-based imputation (Khotimah et al., 2019). This comparison aims to highlight the impact of the proposed method on improving recommendation accuracy through better criteria rating completion.

## RESULTS

The proposed method is compared to baseline methods on three public datasets to demonstrate its impact on the MCCF performance, which consists of criteria rating prediction, overall rating prediction, and recommended item prediction.

### Criteria rating prediction

The first experiment was conducted to evaluate the impact of RNB imputation on criteria rating prediction. The experimental results are presented in Table 3. As shown in Table 3, RNB consistently improves the performance of MCCF in predicting criteria ratings, as indicated by the lowest average MAE across most criteria and datasets. Furthermore, a statistical significance test was conducted to strengthen the reliability of the findings. Tables 4 and 5 summarize the results of the significance tests, comparing the MAE values of the criteria rating prediction between MCCF with RNB imputation and MCCF with baseline imputations.

Based on Table 4, the Analysis of Variance (ANOVA) test revealed a statistically significant difference in performance among the compared methods. A post-hoc analysis using Tukey's Honest Significant Difference (HSD) test was then conducted to identify which methods differed significantly. As shown in Table 5, for the TA and BA datasets, the performance improvements achieved by RNB were statistically significant compared to all baseline methods. Meanwhile, for the YM dataset, RNB achieved the lowest MAE, and the improvement was statistically significant when compared with NI. However, although the difference was not statistically significant when compared with MEAN and NB ($p > 0.05$), RNB still consistently demonstrated superior performance. These findings confirm the positive effect of RNB on enhancing the performance of criteria rating prediction within the MCCF framework.

**Table 3. Average MAE of criteria rating prediction by imputation method**

| Dataset | Criteria | Average MAE | | | |
|---|---|---|---|---|---|
| | | NI | Mean | NB | RNB |
| TA | Service | 0.4246 | 0.3555 | 0.3934 | **0.3537** |
| | Cleanliness | 0.3725 | 0.2866 | 0.3193 | **0.2864** |
| | Value | 0.4230 | 0.3665 | 0.3929 | **0.3659** |
| | Sleep Quality | 0.8288 | 0.3552 | 0.3757 | **0.3384** |
| | Rooms | 0.4972 | 0.3422 | 0.3718 | **0.3396** |
| | Location | 0.5171 | **0.2646** | 0.3364 | 0.2675 |
| YM | Story | 0.8924 | 0.7416 | **0.7257** | 0.7263 |
| | Acting | 2.0228 | 0.7483 | 0.6863 | **0.6834** |
| | Direction | 1.1563 | 0.7541 | **0.7187** | 0.7191 |
| | Visuals | 1.6338 | 0.7486 | 0.7078 | **0.7036** |
| BA | Appearance | 1.2286 | **0.3047** | 0.3195 | 0.3061 |
| | Aroma | 0.5926 | 0.4097 | **0.4085** | 0.4086 |
| | Palate | 1.3869 | 0.4592 | 0.4563 | **0.4468** |
| | Taste | 1.4052 | 0.5167 | 0.5011 | **0.4870** |

**Table 4. ANOVA of criteria rating prediction**

| Dataset | Methods | Count | Mean | Variance | F | $p$ | Remark |
|---|---|---|---|---|---|---|---|
| TA | NI | 5 | 0.5105 | 6.72E-07 | 25370.33 | 1.27E-29 | Significant |
| | MEAN | 5 | 0.3284 | 1.55E-06 | | | |
| | NB | 5 | 0.3649 | 2.67E-06 | | | |
| | RNB | 5 | 0.3253 | 1.12E-06 | | | |
| YM | NI | 5 | 1.4467 | 0.0018 | 651.67 | 6.27E-17 | Significant |
| | MEAN | 5 | 0.7606 | 0.0008 | | | |
| | NB | 5 | 0.7222 | 0.0007 | | | |
| | RNB | 5 | 0.7205 | 0.0007 | | | |
| BA | NI | 5 | 1.1533 | 4.70E-06 | 214608.70 | 4.86E-37 | Significant |
| | MEAN | 5 | 0.4226 | 4.58E-06 | | | |
| | NB | 5 | 0.4214 | 1.43E-06 | | | |
| | RNB | 5 | 0.4121 | 1.86E-06 | | | |

**Table 5. Tukey HSD post hoc test of criteria rating prediction**

| Dataset | Methods | Mean Diff. | p-adj | Remark |
|---|---|---|---|---|
| TA | MEAN – NB | 0.0365 | < 0.001 | Significant |
| | MEAN – NI | 0.1821 | < 0.001 | Significant |
| | MEAN – RNB | −0.0032 | 0.0041 | **Significant** |
| | NB – NI | 0.1456 | < 0.001 | Significant |
| | NB – RNB | −0.0396 | < 0.001 | **Significant** |
| | NI – RNB | −0.1853 | < 0.001 | **Significant** |
| YM | MEAN – NB | −0.0384 | 0.2489 | Not significant |
| | MEAN – NI | 0.686 | < 0.001 | Significant |
| | MEAN – RNB | −0.0401 | 0.2176 | **Not significant** |
| | NB – NI | 0.7245 | < 0.001 | Significant |
| | NB – RNB | −0.0017 | 0.9998 | **Not significant** |
| | NI – RNB | −0.7262 | < 0.001 | **Significant** |
| BA | MEAN – NB | −0.0012 | 0.7222 | Not significant |
| | MEAN – NI | 0.7308 | < 0.001 | Significant |
| | MEAN – RNB | −0.0104 | < 0.001 | **Significant** |
| | NB – NI | 0.732 | < 0.001 | Significant |
| | NB – RNB | −0.0092 | < 0.001 | **Significant** |
| | NI – RNB | −0.7412 | < 0.001 | **Significant** |

## Recommended item prediction

The second experiment was conducted to evaluate the impact of RNB imputation on overall rating prediction and recommended item prediction. The experimental results are presented in Tables 6 and 9, respectively. Moreover, a statistical significance test using ANOVA (Tables 7 and 10) and a post-hoc analysis using Tukey's HSD (Tables 8 and 11) was conducted to confirm the validity of the findings. They present a comparative analysis of prediction accuracy, highlighting the results of MCCF with RNB imputation versus MCCF with baseline imputation.

Table 6 shows that RNB achieves the lowest MAE across all datasets, outperforming baseline methods in predicting overall ratings. The ANOVA test in Table 7 revealed a statistically significant difference in performance among the compared methods. According to the Tukey HSD post-hoc analysis in Table 8, there are significant differences in MAE between most imputation methods. Notably, RNB consistently outperforms other methods across datasets, with statistically significant improvements compared to NI and MEAN. However, in the YM and BA datasets, the differences between RNB and NB were not statistically significant. It indicates comparable performance levels, even though RNB maintains the lowest MAE values.

**Table 6. Average MAE of overall rating prediction by imputation method**

| Dataset | Average MAE | | | |
|---|---|---|---|---|
| | NI | Mean | NB | RNB |
| TA | 0.8397 | 0.5239 | 0.5809 | **0.4781** |
| YM | 1.2986 | 0.7464 | 0.7051 | **0.7038** |
| BA | 1.2158 | 0.5426 | 0.5313 | **0.5308** |

**Table 7. ANOVA of overall rating prediction**

| Dataset | Methods | Count | Mean | Variance | F | $p$ | Remark |
|---|---|---|---|---|---|---|---|
| TA | NI | 5 | 0.8397 | 1.98E-06 | 38569.70 | 4.46E-31 | Significant |
| | MEAN | 5 | 0.5239 | 2.67E-06 | | | |
| | NB | 5 | 0.5809 | 6.44E-06 | | | |
| | RNB | 5 | 0.4781 | 2.45E-06 | | | |
| YM | NI | 5 | 1.2986 | 3.65E-05 | 25242.11 | 1.32E-29 | Significant |
| | MEAN | 5 | 0.7464 | 1.22E-05 | | | |
| | NB | 5 | 0.7051 | 1.38E-05 | | | |
| | RNB | 5 | 0.7038 | 4.51E-06 | | | |
| BA | NI | 5 | 1.2158 | 7.95E-06 | 141406.46 | 1.37E-35 | Significant |
| | MEAN | 5 | 0.5426 | 4.21E-06 | | | |
| | NB | 5 | 0.5313 | 2.99E-06 | | | |
| | RNB | 5 | 0.5307 | 1.25E-06 | | | |

**Table 8. Tukey HSD post hoc test of overall rating prediction**

| Dataset | Methods | Mean diff. | p-adj | Remark |
|---|---|---|---|---|
| TA | MEAN – NB | 0.057 | < 0.001 | Significant |
| | MEAN – NI | 0.3157 | < 0.001 | Significant |
| | MEAN – RNB | -0.0458 | < 0.001 | **Significant** |
| | NB – NI | 0.2588 | < 0.001 | Significant |
| | NB – RNB | -0.1028 | < 0.001 | **Significant** |
| | NI – RNB | -0.3616 | < 0.001 | **Significant** |
| YM | MEAN – NB | -0.0413 | < 0.001 | Significant |
| | MEAN – NI | 0.5522 | < 0.001 | Significant |
| | MEAN – RNB | -0.0426 | < 0.001 | **Significant** |
| | NB – NI | 0.5935 | < 0.001 | Significant |
| | NB – RNB | -0.0013 | 0.9564 | **Not Significant** |
| | NI – RNB | -0.5948 | < 0.001 | **Significant** |
| BA | MEAN – NB | -0.0113 | < 0.001 | Significant |
| | MEAN – NI | 0.6731 | < 0.001 | Significant |
| | MEAN – RNB | -0.0119 | < 0.001 | **Significant** |
| | NB – NI | 0.6844 | < 0.001 | Significant |
| | NB – RNB | -0.0005 | 0.9739 | **Not Significant** |
| | NI – RNB | -0.685 | < 0.001 | **Significant** |

In the task of recommended item prediction (Table 9), RNB consistently delivered the highest performance in all datasets for both moderately (threshold = 3) and highly (threshold = 4) recommended items. The ANOVA results in Table 10 highlighted a statistically significant performance difference among the methods. Table 11 presents the Tukey HSD post-hoc analysis. For the moderately recommended items, the results indicate that RNB significantly outperformed NI across all datasets. In the TA dataset, RNB was consistently statistically significant across all methods. It clearly confirms the performance differences between RNB and other methods. In the YM and BA datasets, the improvement was not statistically significant when compared with NB or MEAN. However, RNB consistently achieved the lowest MAE, indicating its comparable performance. For the highly recommended items, a similar trend was observed. RNB continued to demonstrate the best MAE performance across all datasets. The TA and BA datasets highlight remarkable advancements in RNB. The exception occurred in the YM dataset between NB and RNB, where the difference was not significant. However, the performance between these two methods is still comparable. Overall, these results reinforce the robustness of RNB in maintaining low MAE across various datasets and thresholds, with consistently significant improvements over the NI and competitive performance relative to MEAN and NB. These findings highlight the promising potential of RNB in enhancing the effectiveness of MCCF-based recommendation systems.

**Table 9. Average Macro F1-scores of recommended item prediction by imputation method**

| Threshold | Dataset | Average macro F1-scores | | | |
|---|---|---|---|---|---|
| | | NI | Mean | NB | RNB |
| 3 | TA | 0.640 | 0.700 | 0.730 | **0.758** |
| | YM | 0.300 | 0.728 | **0.732** | **0.732** |
| | BA | 0.390 | 0.682 | **0.700** | **0.700** |
| 4 | TA | 0.550 | 0.730 | 0.718 | **0.750** |
| | YM | 0.440 | 0.474 | 0.600 | **0.604** |
| | BA | 0.310 | 0.470 | 0.474 | **0.480** |

**Table 10. ANOVA of recommended item prediction**

| Dataset | Methods | Count | Mean | Variance | F | *p* | Remark |
|---|---|---|---|---|---|---|---|
| **Threshold = 3** | | | | | | | |
| TA | NI | 5 | 0.64 | 0 | 2556 | 1.18E-21 | Significant |
| | MEAN | 5 | 0.7 | 0 | | | |
| | NB | 5 | 0.73 | 0 | | | |
| | RNB | 5 | 0.758 | 0.00002 | | | |
| YM | NI | 5 | 0.3 | 0 | 15457.33 | 6.69E-28 | Significant |
| | MEAN | 5 | 0.728 | 0.00002 | | | |
| | NB | 5 | 0.732 | 0.00002 | | | |
| | RNB | 5 | 0.732 | 0.00002 | | | |
| BA | NI | 5 | 1.2158 | 7.95E-06 | 23176 | 2.62E-29 | Significant |
| | MEAN | 5 | 0.5426 | 4.21E-06 | | | |
| | NB | 5 | 0.5313 | 2.99E-06 | | | |
| | RNB | 5 | 0.5308 | 1.25E-06 | | | |
| **Threshold = 4** | | | | | | | |
| TA | NI | 5 | 0.55 | 0 | 8516 | 7.86E-26 | Significant |
| | MEAN | 5 | 0.73 | 0 | | | |

| Dataset | Methods | Count | Mean | Variance | F | *p* | Remark |
|---|---|---|---|---|---|---|---|
| | NB | 5 | 0.718 | 0.00002 | | | |
| | RNB | 5 | 0.75 | 0 | | | |
| YM | NI | 5 | 0.44 | 3.85E-33 | 2401.22 | 1.94E-21 | Significant |
| | MEAN | 5 | 0.474 | 3.00E-05 | | | |
| | NB | 5 | 0.6 | 0 | | | |
| | RNB | 5 | 0.604 | 3.00E-05 | | | |
| BA | NI | 5 | 0.31 | 0 | 4530.44 | 1.22E-23 | Significant |
| | MEAN | 5 | 0.47 | 3.85E-33 | | | |
| | NB | 5 | 0.474 | 3E-05 | | | |
| | RNB | 5 | 0.48 | 0 | | | |

**Table 11. Tukey HSD post hoc test of recommended item prediction**

| Dataset | Methods | Mean diff. | p-adj | Remark |
|---|---|---|---|---|
| **Threshold = 3** | | | | |
| TA | MEAN – NB | 0.03 | < 0.001 | Significant |
| | MEAN – NI | -0.06 | < 0.001 | Significant |
| | MEAN – RNB | 0.058 | < 0.001 | **Significant** |
| | NB – NI | -0.09 | < 0.001 | Significant |
| | NB – RNB | 0.028 | < 0.001 | **Significant** |
| | NI – RNB | 0.118 | < 0.001 | **Significant** |
| YM | MEAN – NB | 0.004 | 0.3891 | Not Significant |
| | MEAN – NI | -0.428 | < 0.001 | Significant |
| | MEAN – RNB | 0.004 | 0.3891 | **Not Significant** |
| | NB – NI | -0.432 | < 0.001 | Significant |
| | NB – RNB | 0 | 1 | **Not Significant** |
| | NI – RNB | 0.432 | < 0.001 | **Significant** |
| BA | MEAN – NB | 0.018 | < 0.001 | Significant |
| | MEAN – NI | -0.292 | < 0.001 | Significant |
| | MEAN – RNB | 0.018 | < 0.001 | **Significant** |
| | NB – NI | -0.31 | < 0.001 | Significant |
| | NB – RNB | 0 | 1 | **Not Significant** |
| | NI – RNB | 0.31 | < 0.001 | **Significant** |
| **Threshold = 4** | | | | |
| TA | MEAN – NB | -0.012 | < 0.001 | Significant |
| | MEAN – NI | -0.18 | < 0.001 | Significant |
| | MEAN – RNB | 0.02 | < 0.001 | **Significant** |
| | NB – NI | -0.168 | < 0.001 | Significant |
| | NB – RNB | 0.032 | < 0.001 | **Significant** |
| | NI – RNB | 0.2 | < 0.001 | **Significant** |
| YM | MEAN – NB | 0.126 | < 0.001 | Significant |

| Dataset | Methods | Mean diff. | p-adj | Remark |
|---|---|---|---|---|
| | MEAN – NI | -0.034 | < 0.001 | Significant |
| | MEAN – RNB | 0.13 | < 0.001 | **Significant** |
| | NB – NI | -0.16 | < 0.001 | Significant |
| | NB – RNB | 0.004 | 0.3891 | **Not Significant** |
| | NI – RNB | 0.164 | < 0.001 | **Significant** |
| BA | MEAN – NB | 0.004 | 0.1373 | Not Significant |
| | MEAN – NI | -0.16 | < 0.001 | Significant |
| | MEAN – RNB | 0.01 | 0.0002 | **Significant** |
| | NB – NI | -0.164 | < 0.001 | Significant |
| | NB – RNB | 0.006 | 0.0152 | **Significant** |
| | NI – RNB | 0.17 | < 0.001 | **Significant** |

## FINDINGS

*RQ: How does RNB-based imputation improve MCRS performance?*

Answering the research question, the main findings show RNB's consistent superiority over baseline methods in enhancing the performance of MCRS, especially MCCF, for all datasets. A notable improvement has been achieved in predicting criteria ratings, overall ratings, and recommended items. The most significant improvements are demonstrated in datasets with more criteria, stronger inter-criteria correlations, and larger sizes, such as the TA dataset. Furthermore, RNB is shown to improve recommendation accuracy, either for moderately or highly recommended items. Overall, these findings indicate that RNB is an effective and adaptive approach to address the partial preference problem in MCRS. RNB is also capable of providing practical benefits in the context of recommending more relevant items to users.

## DISCUSSION

The experimental results clearly demonstrate the effectiveness of the proposed RNB method in enhancing the performance of MCCF. By addressing the issue of missing criteria ratings through a structured imputation process, RNB enables more complete and informative input for subsequent rating and recommendation predictions. Unlike traditional NB imputations that treat criteria independently, RNB introduces a sequential dependency structure that incrementally expands the feature set as more criteria are imputed. This recurrent mechanism provides a technically grounded way to capture inter-criteria dependencies, which aligns with prior findings that multi-criteria relationships can significantly influence rating behaviour (Zheng, 2017). The consistent improvements observed across all datasets in both criteria rating prediction (Table 3) and overall rating prediction (Table 6) suggest that the use of more reliable and complete data in MCCF pipelines contributes significantly to reducing prediction error.

Moreover, the results of recommended item prediction (Table 9) further highlight the practical relevance of the proposed approach. Although some baseline methods (e.g., Mean and NB imputation) occasionally perform comparably under specific settings, RNB consistently yields the highest Macro F1-score for both moderately and highly recommended items. This indicates that RNB not only improves prediction accuracy but also enhances the model's ability to distinguish between relevant and irrelevant items, which is crucial for real-world recommendation applications.

Performance differences across datasets also provide insight into how RNB interacts with dataset characteristics. The TA dataset exhibited the most remarkable improvements. This could be due to three main factors: (1) a greater number of criteria, providing richer contextual information for imputation; (2) a higher inter-criteria correlation, as shown in Figure 5, which enhances the predictive power of the recurrent structure in RNB using already imputed or known values; and (3) a larger dataset, allowing for more reliable estimation of conditional probabilities, which leads to more accurate imputations and consequently, better recommendation performance. These results confirm that stronger correlations amplify the benefits of dependency-aware imputation (Fan et al., 2023). Overall, these findings emphasize the relationship between the effectiveness of RNB and the dataset characteristics, including the number of criteria, inter-criteria correlation, and dataset size.

The ordering strategy used in RNB, based on the lowest sparsity first, also proved to be beneficial. By predicting criteria with more complete data early on, the model builds a more stable foundation for subsequent imputations, reducing the risk of error propagation. While this approach may not always prioritize the most informative (i.e., highly correlated) criteria, it ensures that the initial predictions are statistically reliable, which is particularly important when working with sparse data (Choi et al., 2023).

Despite these strengths, RNB has some limitations. Its performance relies on the assumption that inter-criteria dependencies can be captured sequentially, which may not be well-applied in all domains. Furthermore, the optimal ordering of criteria may vary across datasets, requiring dataset-specific analysis. Suboptimal criteria ordering has the potential to propagate imputation errors across successive steps. Consequently, it may adversely affect the overall model performance. Future work may explore dynamic or hybrid ordering strategies that consider both sparsity and correlation simultaneously. It is also worthwhile to leverage more advanced machine learning techniques, which can handle more complex or nonlinear dependencies, such as combining RNB with probabilistic graphical models or neural architectures.

In summary, the proposed RNB method offers a simple yet powerful enhancement to existing MCCF frameworks through dependency-aware sequential imputation. The method delivers consistent improvements across datasets and enhances recommendation accuracy, particularly in datasets that contain rich criteria information and exhibit highly correlated criteria. These results confirm that preprocessing strategies targeting missing value imputation are an essential component of effective MCRS.

## CONCLUSION

This study emphasizes the important role of data preprocessing, particularly missing value imputation, in enhancing the quality of recommendations in MCRS. To address the issue of incomplete criteria-level ratings, imputation is required to improve data completeness. Baseline imputation methods treat criteria independently, despite real-world user preferences often exhibiting semantic correlations. To address this issue, we introduced RNB, a novel preprocessing method applied prior to the MCRS framework. RNB introduces a sequential dependency structure that provides a technically grounded way to capture inter-criteria dependencies.

Answering the research question, experiments were conducted on three real-world datasets: TA, YM, and BA. Experimental results demonstrated that RNB consistently outperformed baseline imputation methods (no imputation, mean, and NB) in predicting criteria ratings, overall ratings, and recommended items. These results highlight the promising potential of RNB in improving the overall effectiveness of MCRS. By leveraging inter-criteria dependencies in a sequential manner, RNB effectively enriches the input space for MCRS, enabling more accurate recommendations. Such improvements have meaningful practical implications, including elevated user satisfaction, improved user experience, increased trust in digital platforms, and more informed decision-making. This makes RNB particularly valuable for real-world MCRS applications, such as hotel, product, or movie recommender systems, where incomplete criteria-level ratings are common. However, the degree of improvement varied across datasets, indicating that RNB's performance is influenced by key dataset characteristics such as the number of criteria, inter-criteria correlations, and dataset size. This contextualizes the limitations of baseline methods, which assume independence among criteria or rely on overly simplistic statistical assumptions that fail to exploit richer dependencies.

Despite its advantage, RNB relies on the assumption that inter-criteria dependencies can be captured sequentially. Moreover, the optimal ordering of criteria may vary across datasets. Inaccurate ordering may propagate errors to subsequent steps. Thus, dataset-specific analysis is required to determine an appropriate ordering strategy and to ensure the robustness of the method. Future work may explore adaptive ordering strategies that simultaneously consider sparsity and correlation, as well as incorporate advanced machine learning techniques capable of modeling complex or nonlinear relationships among criteria. These directions hold promise for further enhancing the robustness and generalizability of preprocessing approaches in MCRS.

## REFERENCES

Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, *22*(3), 48–55. https://doi.org/10.1109/MIS.2007.58

Afzal, I., Yilmazel, B., & Kaleli, C. (2024). An approach for multi-context-aware multi-criteria recommender systems based on deep learning. *IEEE Access*, *12*, 99936–99948. https://doi.org/10.1109/ACCESS.2024.3428630

Alam, S., Ayub, M. S., Arora, S., & Khan, M. A. (2023). An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity. *Decision Analytics Journal*, *9*, 100341. https://doi.org/10.1016/j.dajour.2023.100341

Al-Ghuribi, S. M., & Mohd Noah, S. A. (2019). Multi-criteria review-based recommender system – The state of the art. *IEEE Access*, *7*, 169446–169468. https://doi.org/10.1109/ACCESS.2019.2954861

Bag, S., Kumar, S. K., & Tiwari, M. K. (2019). An efficient recommendation generation using relevant Jaccard similarity. *Information Sciences*, *483*, 53–64. https://doi.org/10.1016/j.ins.2019.01.023

Choi, S.-M., Lee, D., Jang, K., Park, C., & Lee, S. (2023). Improving data sparsity in recommender systems using matrix regeneration with item features. *Mathematics*, *11*(2), 292. https://doi.org/10.3390/math11020292

Dehbozorgi, N., Kunuku, M. T., & Pouriyeh, S. (2024). Personalized pedagogy through a LLM-based recommender system. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Artificial intelligence in education. posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky* (pp. 63–70). Springer. https://doi.org/10.1007/978-3-031-64312-5_8

Du, J., Hew, K. F., & Zhang, L. (2025). Designing a recommender system to promote self-regulated learning in online contexts: A design-based study. *Education and Information Technologies*, *30*(2), 2389–2428. https://doi.org/10.1007/s10639-024-12867-w

Fan, G., Zhang, C., Chen, J., Li, P., Li, Y., & Leung, V. C. M. (2023). Improving rating prediction in multi-criteria recommender systems via a collective factor model. *IEEE Transactions on Network Science and Engineering*, *10*(6), 3633–3643. https://doi.org/10.1109/TNSE.2023.3270910

Hong, M., & Jung, J. J. (2021a). ClustPTF: Clustering-based parallel tensor factorization for the diverse multi-criteria recommendation. *Electronic Commerce Research and Applications*, *47*, 101041. https://doi.org/10.1016/j.elerap.2021.101041

Hong, M., & Jung, J. J. (2021b). Hypothetical tensor-based multi-criteria recommender system for new users with partial preferences. *Computer Science and Information Systems*, *18*(1), 285–301. https://doi.org/10.2298/CSIS200531056H

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

Jannach, D., Karakaya, Z., & Gedikli, F. (2012). Accuracy improvements for multi-criteria recommender systems. *Proceedings of the 13th ACM Conference on Electronic Commerce* (pp. 674–689). Association for Computing Machinery. https://doi.org/10.1145/2229012.2229065

Keikhosrokiani, P., & Fye, G. M. (2024). A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings. *Multimedia Tools and Applications*, *83*(15), 45315–45344. https://doi.org/10.1007/s11042-023-17321-6

Khotimah, B. K., Miswanto, & Suprajitno, H. (2019). Modeling naïve Bayes imputation classification for missing data. *IOP Conference Series: Earth and Environmental Science*, *243*, 012111. https://doi.org/10.1088/1755-1315/243/1/012111

Kitson, N. K., Constantinou, A. C., Guo, Z., Liu, Y., & Chobtham, K. (2023). A survey of Bayesian network structure learning. *Artificial Intelligence Review*, *56*(8), 8721–8814. https://doi.org/10.1007/s10462-022-10351-w

Le, Q.-H., Mau, T. N., Tansuchat, R., & Huynh, V.-N. (2022). A multi-criteria collaborative filtering approach using deep learning and Dempster-Shafer theory for hotel recommendations. *IEEE Access*, *10*, 37281–37293. https://doi.org/10.1109/ACCESS.2022.3165310

McAuley, J., & Leskovec, J. (2013). From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. *Proceedings of the 22nd International Conference on World Wide Web* (pp. 897–908). Association for Computing Machinery. https://doi.org/10.1145/2488388.2488466

McAuley, J., Leskovec, J., & Jurafsky, D. (2012). Learning attitudes and attributes from multi-aspect reviews. *Proceedings of the IEEE 12th International Conference on Data Mining*, *Brussels, Belgium,* 1020–1025. https://doi.org/10.1109/ICDM.2012.110

Nassar, N., Jafar, A., & Rahhal, Y. (2020a). A novel deep multi-criteria collaborative filtering model for recommendation system. *Knowledge-Based Systems*, *187*, 104811. https://doi.org/10.1016/j.knosys.2019.06.019

Nassar, N., Jafar, A., & Rahhal, Y. (2020b). Multi-criteria collaborative filtering recommender by fusing deep neural network and matrix factorization. *Journal of Big Data*, *7*, Article 34. https://doi.org/10.1186/s40537-020-00309-6

Opitz, J., & Burst, S. (2021). *Macro F1 and macro F1*. arXiv. http://arxiv.org/abs/1911.03347

Reddy, S. V., Krishna, V. R., Sapkal, R. J., Dhanke, J., Waghmare, S. P., & Kumar, K. (2024). A cutting-edge artificial intelligence paradigm for entertainment-infused music recommendations. *Entertainment Computing*, *51*, 100717. https://doi.org/10.1016/j.entcom.2024.100717

Rismala, R., Maulidevi, N. U., & Surendro, K. (2023). Multi-attribute BERT for preferences completion in multi-criteria recommender system. *Proceedings of the 12th International Conference on Software and Computer Applications* (pp. 315–320). Association for Computing Machinery. https://doi.org/10.1145/3587828.3587875

Rismala, R., Maulidevi, N. U., & Surendro, K. (2024a). Personalized neural network-based aggregation function in multi-criteria collaborative filtering. *Journal of King Saud University – Computer and Information Sciences*, *36*(1), 101922. https://doi.org/10.1016/j.jksuci.2024.101922

Rismala, R., Maulidevi, N. U., & Surendro, K. (2024b). Personalized neural network-based aggregation function with partial preference handling in multi-criteria collaborative filtering. *Proceedings of the 11th International Conference on Advanced Informatics: Concept, Theory and Application*, *Singapore,* 1–5. https://doi.org/10.1109/ICAICTA63815.2024.10763316

Sahu, S., Kumar, R., MohdShafi, P., Shafi, J., Kim, S., & Ijaz, M. F. (2022). A hybrid recommendation system of upcoming movies using sentiment analysis of YouTube trailer reviews. *Mathematics*, *10*(9), 1568. https://doi.org/10.3390/math10091568

Shambour, Q. (2021). A deep learning based algorithm for multi-criteria recommender systems. *Knowledge-Based Systems*, *211*, 106545. https://doi.org/10.1016/j.knosys.2020.106545

Shambour, Q., Abualhaj, M., Abu-Shareha, A., & Kharma, Q. (2024). Personalized tourism recommendations: Leveraging user preferences and trust network. *Interdisciplinary Journal of Information, Knowledge, and Management*, *19*, 17. https://doi.org/10.28945/5329

Shambour, Q., Al-Zyoud, M., Abu-Shareha, A., & Abualhaj, M. (2023). Medicine recommender system based on semantic and multi-criteria filtering. *Interdisciplinary Journal of Information, Knowledge, and Management*, *18*, 435–457. https://doi.org/10.28945/5172

Shankar, A., Perumal, P., Subramanian, M., Ramu, N., Natesan, D., Kulkarni, V. R., & Stephan, T. (2024). An intelligent recommendation system in e-commerce using ensemble learning. *Multimedia Tools and Applications*, *83*(16), 48521–48537. https://doi.org/10.1007/s11042-023-17415-1

Shaw, G. K., Chattopadhyay, R., Paul, A., Rounik, K., Kumar, S., Saha, B., & Indu, P. (2024). Entertainment recommendation and rating system based on emotions. In P. Raj, A. Rocha, S. P. Singh, P. K. Dutta, & B. Sundaravadivazhagan (Eds.), *Building embodied AI systems: The agents, the architecture principles, challenges, and application domains* (pp. 443–462). Springer. https://doi.org/10.1007/978-3-031-68256-8_20

Wang, F., Pan, W., & Chen, L. (2013). Recommendation for new users with partial preferences by integrating product reviews with static specifications. In S. Carberry, S. Weibelzahl, A. Micarelli, & G. Semeraro (Eds.), *User modeling, adaptation, and personalization* (pp. 281–288). Springer. https://doi.org/10.1007/978-3-642-38844-6_24

Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data: A rating regression approach. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 783–792). Association for Computing Machinery. https://doi.org/10.1145/1835804.1835903

Wang, H., Lu, Y., & Zhai, C. (2011). Latent aspect rating analysis without aspect keyword supervision. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 618–626). Association for Computing Machinery. https://doi.org/10.1145/2020408.2020505

Yashudas, A., Gupta, D., Prashant, G. C., Dua, A., AlQahtani, D., & Reddy, A. S. K. (2024). DEEP-CARDIO: Recommendation system for cardiovascular disease prediction using IoT network. *IEEE Sensors Journal*, *24*(9), 14539–14547. https://doi.org/10.1109/JSEN.2024.3373429

Zhang, K., Liu, X., Wang, W., & Li, J. (2021). Multi-criteria recommender system based on social relationships and criteria preferences. *Expert Systems with Applications*, *176*, 114868. https://doi.org/10.1016/j.eswa.2021.114868

Zheng, Y. (2017). Criteria Chains: A novel multi-criteria recommendation approach. *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 29–33). Association for Computing Machinery. https://doi.org/10.1145/3025171.3025215

Zhuang, Y., & Kim, J. (2021). A BERT-based multi-criteria recommender system for hotel promotion management. *Sustainability*, *13*(14), 8039. https://doi.org/10.3390/su13148039

## AUTHORS

**Rita Rismala** received a Doctor of Electrical Engineering and Informatics degree from Bandung Institute of Technology, Indonesia, in 2024. She also received bachelor's and master's degrees in Informatics Engineering from Telkom Institute of Technology (now Telkom University), Bandung, Indonesia, in 2010 and 2013. Currently, she is a senior lecturer in the School of Computing, Telkom University. Her research interests include recommender systems, machine learning, and artificial intelligence.

**Untari Novia Wisesty** received a Doctor of Electrical Engineering and Informatics degree from Bandung Institute of Technology, Indonesia, in 2023. She also received bachelor's and master's degrees in Informatics Engineering from Telkom Institute of Technology (now Telkom University), Bandung, Indonesia, in 2010 and 2012. Since December 2024, she has been an associate professor of Artificial Intelligence at Telkom University. Her research interests include machine learning, deep learning, artificial intelligence, bioinformatics, and biomedical engineering.

**Febryanti Sthevanie** is an Indonesian academic and lecturer in the field of Informatics and Computer Science, specializing in computer vision, image processing, pattern recognition, object detection, and biometrics. She holds the academic titles S.T. and M.T. (Bachelor of Technology and Master of Technology) and is affiliated with Telkom University in Bandung, Indonesia, where she teaches in the School of Computing within the Faculty of Informatics.