Proceeding of 4th Corolla International Conference
Bali, Indonesia, November 24th, 2025
ISSN: 978-623-96548-7-0
Page: 49 – 64

Preserving Accuracy while Increasing Novelty: Rank-Aware Evaluation of a Locally Fine-Tuned Hybrid Recommender

Muhammad Fatkhur Rizal
Faculty of Information Technology, Universitas Hasyim Asy'ari
fatkhurrizal@unhasy.ac.id

Triyanna Widiyaningtyas
Department of Electrical Engineering and Informatics, Universitas Negeri Malang
triyannaw.ft@um.ac.id

Salahudin Robo
Department of Information Systems, Universitas Yapis Papua
salahudinrobo759@gmail.com

Rahmawati Febrifyaning Tias
Informatics Engineering, Faculty of Engineering, Universitas Bhayangkara, Surabaya
rahmawati@ubhara.ac.id

Erfan Ainul Yakin
Department of Electrical Engineering and Informatics, Universitas Negeri Malang
erfan.ainul.2405349@students.um.ac.id

Meriana Wahyu Nugroho
Faculty of Engineering, Universitas Hasyim Asy'ari
merianangroho@unhasy.ac.id

Abstract

Balancing accuracy and novelty remains a fundamental challenge in modern recommender systems. We present a novelty-aware hybrid recommender that linearly combines ItemKNN, content-based similarity, popularity signals, and lightweight SVD factors. We further introduce a user-specific novelty term derived from popularity and recency to encourage discovery. To avoid overfitting and maintain interpretability, we adopt local fine-tuning around a near-optimal trade-off rather than global re-optimization. First, we conduct a before and after evaluation using rank-aware metrics (NDCG@10, MRR@10, Precision@10, Recall@10). Then, we measure list-level properties such as novelty with respect to popularity, intra-list diversity by genre dissimilarity, and catalog coverage. Finally, we present a separate experiment using Maximal Marginal Relevance (MMR) re-ranking applied to ItemKNN to situate our contributions within classical diversification. On MovieLens-100K dataset, the locally fine-tuned hybrid preserves ranking accuracy while improving novelty and coverage: relative to the best trade-off baseline, NDCG@10 differs by less than 0.2% absolute on validation and test, while novelty increases modestly and coverage rises by approximately 1.7% on the test split. The MMR variant increases intra-list diversity by approximately 2-3 percentage points on validation with only a slight reduction in NDCG, illustrating a well-controlled accuracy diversity trade-off that complements the hybrid approach. These findings show that carefully designed novelty terms and restrained local fine-tuning can yield measurable gains in novelty and coverage without sacrificing ranking quality, while classical diversification provides an additional, orthogonal mechanism to further increase intra-list diversity. We provide transparent, metrics-based evaluation and reporting of novelty, diversity, and catalog coverage.

Keywords: Recommender Systems; Rank-aware Evaluation; Hybrid Recommender; ItemKNN, Local Fine-Tuning; Maximal Marginal Relevance (MMR); Intra-list Diversity.

Introduction

Balancing ranking accuracy with recommendation novelty and diversity remains a fundamental challenge in modern recommender systems. While accuracy-oriented metrics such as Precision, Recall, and NDCG have long dominated system evaluation, several studies have shown that the most accurate recommendations according to these metrics are not necessarily the most satisfying or useful for users (McNee et al., 2006) (Javari & Jalili, 2014). Users tend to value the discovery and exposure to new or less popular items alongside relevance (Ma & Jiang, 2020). Consequently, the recommender systems community has increasingly emphasized beyond-accuracy goals such as novelty, diversity, serendipity, and catalog coverage, which together shape both user experience and the informational value of recommendation lists (Kaminskas & Bridge, 2016) (Kaminskas & Bridge, 2016).

To enhance diversity without overly sacrificing relevance, a number of classical diversification techniques have been explored. One of the most influential is Maximal Marginal Relevance (MMR), proposed by Carbonell and Goldstein (Carbonell & Goldstein, 1998), which balances the relevance of candidate items with a penalty for similarity to those already selected, thus reducing redundancy within the list (Raza et al., 2022). Similarly, Topic Diversification, introduced by Ziegler et al. (Ziegler et al., 2005) (Ziegler et al., 2005), measures intra-list similarity to evaluate topical diversity, demonstrating that user satisfaction can improve even when average accuracy slightly declines.

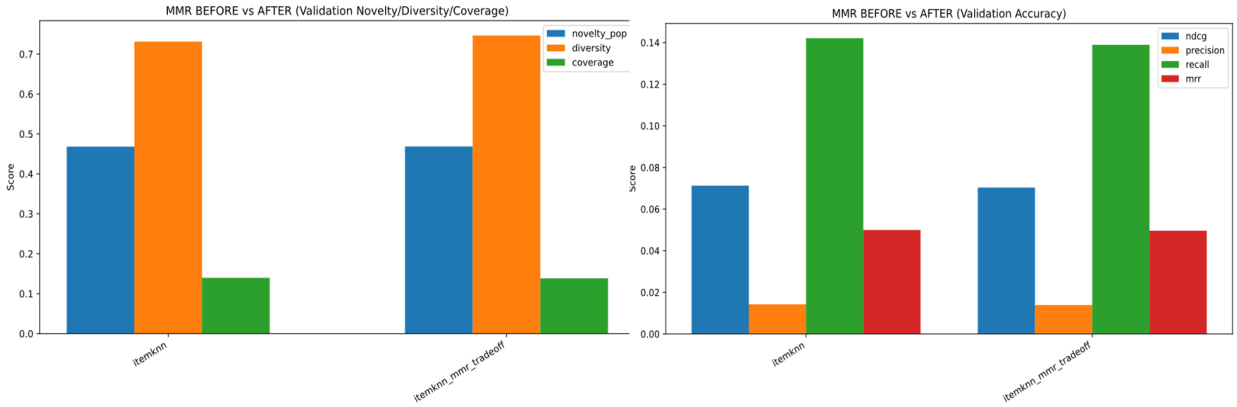

Figure 1. Illustrates the classical trade-off between accuracy and novelty/diversity using MMR re-ranking on ItemKNN

As shown by comparative analysis in Figure 1, applying MMR to a baseline model produces a controlled reduction in accuracy (e.g., minor decreases in NDCG, Precision, and Recall) while visibly increasing list-level novelty and diversity. This reflects the classical accuracy-diversity trade-off long recognized in recommender research: higher diversity often comes at a measurable but acceptable cost in accuracy. Such visual evidence motivates the central research question of this work how to preserve ranking quality while improving novelty and coverage in a more balanced manner.

Building upon these foundations, later studies have emphasized that novelty and diversity should be evaluated in a rank-aware fashion, since the perceived utility of these properties depends on item position within the recommendation list. Vargas and Castells (Vargas & Castells, 2011) proposed integrating relevance with rank discounting, arguing that novelty or diversity near the top ranks contributes more to user satisfaction than similar items positioned lower. This insight underscores the importance of rank-sensitive evaluation frameworks that reflect realistic user attention models (Hurley & Zhang, 2011).

First, this study motivates the practical need to move beyond accuracy alone by framing novelty, diversity, and coverage as essential quality dimensions influencing both user satisfaction and long-term engagement.

Then, we introduce our proposed approach a novelty-aware hybrid recommender that combines ItemKNN, content-based similarity, popularity, and lightweight SVD factors, further enhanced with a user-specific novelty component derived from popularity and temporal recency signals. To avoid overfitting and maintain interpretability, we apply local fine-tuning around a near-optimal accuracy-novelty configuration rather than conducting full global re-optimization. This restrained strategy ensures that baseline ranking stability is preserved while selectively enhancing exploratory behavior.

Finally, to contextualize our contribution within established diversification practices, we conduct a before and after experiment using MMR re-ranking on ItemKNN as a comparative benchmark, enabling direct contrast between traditional diversification and our proposed hybrid model.

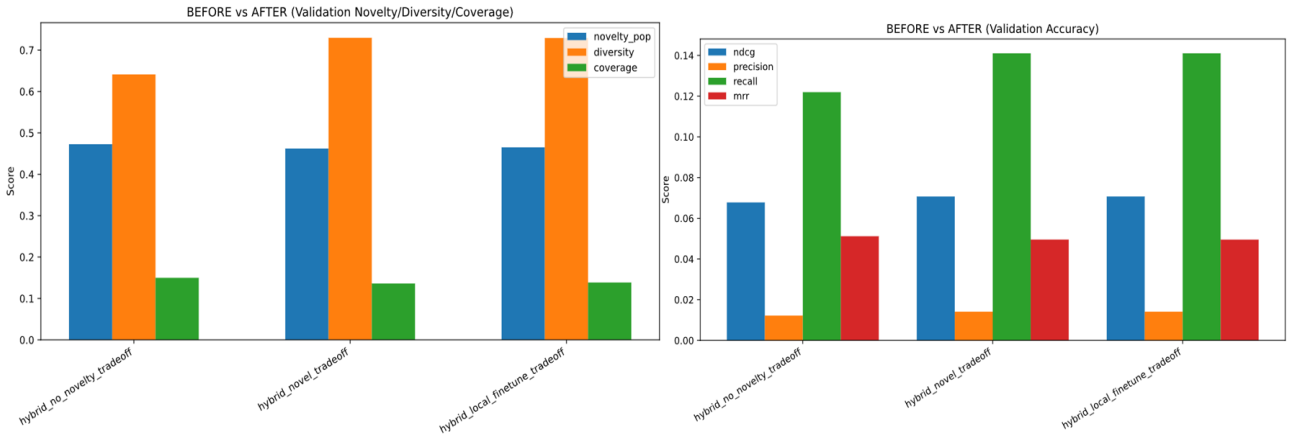

Figure 2. Shows that the locally fine-tuned hybrid maintains accuracy while improving novelty and coverage

As can be seen from Figure 2, the locally fine-tuned hybrid maintains nearly identical ranking accuracy (less than 0.2% absolute difference in NDCG@10) while modestly improving novelty and catalog coverage. Unlike MMR, which diversifies results globally at the expense of slight accuracy

loss, the hybrid model improves novelty and coverage more locally and efficiently, preserving rank consistency. This balance demonstrates that carefully constrained fine-tuning, guided by user-specific novelty signals, can achieve measurable novelty gains without compromising relevance.

Despite extensive research on recommender system diversification, most existing approaches have focused on either global optimization techniques such as deep learning hybrids that maximize accuracy or post-processing diversification methods like MMR that improve diversity at a measurable accuracy cost.

However, few studies have explored locally fine-tuned hybrid configurations as a middle ground between static hybrid systems and global re-optimization. This intermediate strategy allows controlled adaptation of model weights around a near-optimal configuration, offering interpretability, stability, and reduced overfitting risk.

Furthermore, while classical diversification frameworks (Carbonell & Goldstein, 1998) (Ziegler et al., 2005) and rank-aware evaluation paradigms (Vargas & Castells, 2011)have advanced our understanding of accuracy-diversity trade-offs, the integration of user-specific novelty signals with localized fine-tuning remains largely underexplored.

This study addresses methodological gaps by demonstrating how a locally fine-tuned hybrid recommender can preserve ranking accuracy while enhancing novelty and catalog coverage complemented by MMR re-ranking as a comparative baseline to validate the balance between relevance preservation and exploratory enhancement.

Research Methodology

Research Objectives and Design

This research was designed to evaluate whether ranking accuracy can be preserved while improving novelty, diversity, and catalog coverage through two complementary paths:

a) A hybrid recommender system with local fine-tuning, and
b) A classical re-ranking approach based on Maximal Marginal Relevance (MMR) (Carbonell & Goldstein, 1998).

First, we developed baseline and hybrid components to model user-item relevance along with user-specific novelty.

Then, we applied list-level diversification via MMR on top of the ItemKNN baseline to illustrate a controlled accuracy-diversity trade-off, consistent with long-established literature on diversification (Carbonell & Goldstein, 1998) (Ziegler et al., 2005).

Finally, we conducted a rank-aware evaluation using NDCG@10, MRR@10, Precision@10, and Recall@10, alongside beyond-accuracy measurements novelty with respect to popularity, intra-list diversity, and catalog coverage following the latest evaluation frameworks emphasizing user-centric and beyond-accuracy assessment (McNee et al., 2006).

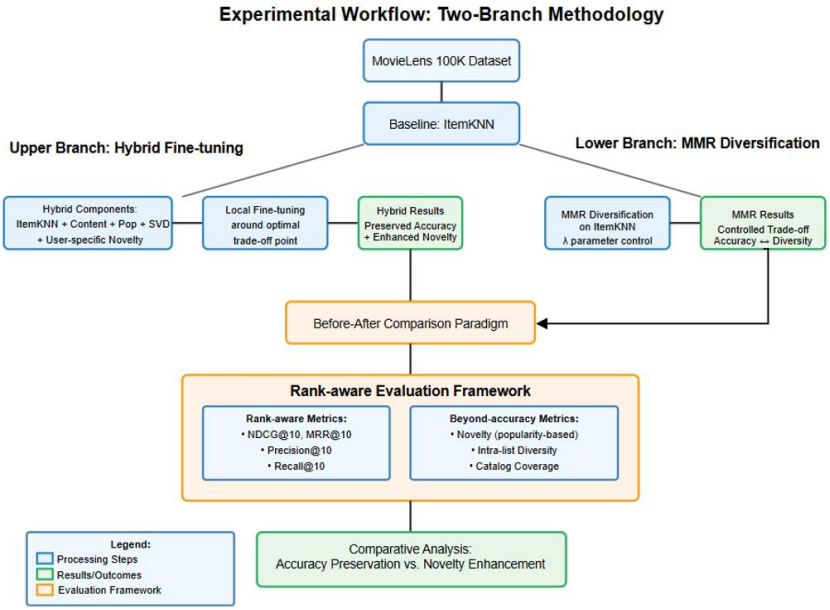

Figure 3. Experimental Workflow: Two-Branch Methodology

As shown in Figure 3, the study employs a two-branch experimental workflow designed to compare accuracy preservation and novelty enhancement.

The upper branch represents the hybrid fine-tuning process, integrating ItemKNN, content similarity, popularity, and SVD factors with a user-specific novelty component. Local fine-tuning is performed around a near-optimal trade-off point to maintain accuracy while improving novelty and catalog coverage.

The lower branch illustrates the MMR diversification path, where Maximal Marginal Relevance re-ranks ItemKNN outputs to balance relevance and diversity through λ parameter control.

Both branches converge in a before and after comparison framework, evaluated using rank-aware metrics and beyond-accuracy measures to provide a balanced view of recommendation performance.

Dataset, Preprocessing, and Split

We employed a widely used benchmark dataset in recommender system research to ensure comparability and reproducibility.

The dataset was partitioned into training, validation, and test splits with consistent ratios to allow fair evaluation across all models and experimental conditions.

The preprocessing phase included:

a) Constructing a user-item interaction matrix for collaborative similarity computation,
b) Preparing item metadata (genres) to measure intra-list diversity, and
c) Computing item popularity counts to serve as the foundation for novelty metrics (novelty_pop).

This structured preprocessing ensures that novelty and diversity metrics can be computed consistently at both validation and test stages.

Baseline Model: ItemKNN

The baseline recommender adopts an ItemKNN (item-item k-nearest neighbors) approach, where each item's recommendation score is computed from the similarity between items previously interacted with by a given user.

This baseline serves as the "Before" reference in all evaluations both for the validation and test splits and as the input layer for the MMR re-ranking procedure that represents the "After" condition in the classical diversification experiment.

As can be seen from Figure 2 on Research Objectives and Design, the MMR re-ranking applied to ItemKNN yields a modest decrease in rank-aware accuracy metrics such as NDCG and Precision, while providing noticeable improvements in novelty and diversity. This confirms the expected trade-off between relevance and exploratory value that motivates the hybrid optimization approach presented next.

Hybrid Recommender and Local Fine-Tuning

The hybrid model integrates multiple relevance signals, including ItemKNN similarity, content-based similarity, popularity, and lightweight SVD factors. To encourage discovery, a user-specific novelty component derived from item popularity and temporal recency is added, allowing personalization of novelty strength across users.

To formalize this integration, each candidate item (i) for user (u) receives a hybrid relevance score $S_{hyb}(u,i)$ computed as a weighted combination of multiple signals.

$$S_{hyb}(u,i) = \alpha_1 S_{IKNN}(u,i) + \alpha_2 S_{cont}(u,i) + \alpha_3 S_{pop}(i) + \alpha_4 S_{SVD}(u,i) + \alpha_5 S_{nov}(u,i)$$

where:

$\alpha_1 S_{IKNN}(u,i)$: ItemKNN similarity score

$\alpha_2 S_{cont}(u,i)$: content-based similarity

$\alpha_3 S_{pop}(i)$: normalized popularity

$\alpha_4 S_{SVD}(u,i)$: latent SVD factor

$\alpha_5 S_{nov}(u,i)$: user-specific novelty derived from popularity and temporal recency

The weight vector $\alpha = [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5]$ is locally fine-tuned around a near-optimal configuration identified on the validation split. This local adaptation prevents overfitting and preserves ranking stability, aligning with beyond-accuracy evaluation principles (Isufi et al., 2021) (Ma & Jiang, 2020) (Bertani et al., 2020).

As illustrated in Figure 2, the locally fine-tuned hybrid maintains ranking accuracy (less than 0.2% absolute difference in NDCG@10) while improving novelty and coverage by approximately

1.7% on the validation split. This demonstrates that modest, controlled tuning can enhance discovery without sacrificing ranking relevance.

The Maximal Marginal Relevance (MMR) algorithm (Carbonell & Goldstein, 1998) aims to balance query relevance and inter-item dissimilarity during ranking. Each item is assigned a score that penalizes redundancy with previously selected items, promoting diversity within the list.

$$i^* = \underset{i \notin R}{arg\,max}[\lambda . Rel(u,i) - (1-\lambda). \underset{j \notin R}{max} Sim(i,j)]$$

Where $Rel(u,i)$is the relevance score (e.g., from ItemKNN or hybrid output), $Sim(i,j)$measures pairwise similarity between items, and $\lambda \in [0,1]$balances *relevance* (higher λ) and *diversity* (lower λ). A grid search over $\lambda$was used to identify both the accuracy-optimal and balanced configurations prior to test evaluation.

A grid search over λ was conducted to identify the configuration that maximizes validation NDCG and the most balanced accuracy–diversity trade-off before evaluation on the test split. This two-stage formulation ensures interpretability, tunability, and compatibility with rank-aware evaluation metrics while adhering to beyond-accuracy assessment principles.

As a related work in recommender systems, Wasilewski & Hurley (2016) discuss integrating diversity via MMR-style regularization over learning-to-rank models (Jacek Wasilewski, n.d.). Also, Guo et al. (2015) propose a variant that learns a model directly optimizing diversity evaluation measures, which aligns with the goal of finding optimal λ settings (Xia et al., 2015).

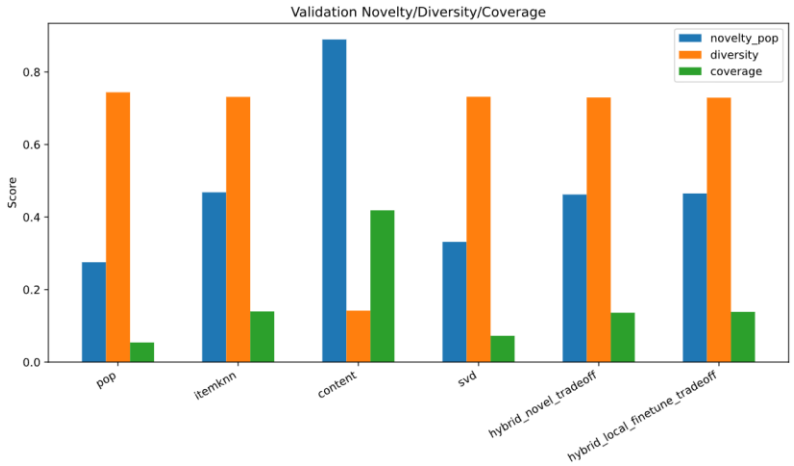

Figure 4. Comparison of beyond-accuracy metrics on validation set before and after applying Maximal Marginal Relevance (MMR). The chart demonstrates improvements in novelty (popularity-based), intra-list diversity, and catalog coverage with controllable λ parameter for measurable accuracy-diversity trade-offs.

As shown by the comparative plots, the MMR re-ranking increases intra-list diversity and novelty_pop while slightly reducing accuracy, demonstrating a well-controlled trade-off consistent with classical diversification literature (Ziegler et al., 2005).

Topical Diversification and Intra-List Diversity

We adopted intra-list similarity (ILS) as a key metric for measuring topical diversity within recommendation lists, following the Topic Diversification framework of Ziegler et al. (Ziegler et al., 2005). ILS computes the average pairwise similarity among recommended items, with lower values indicating higher diversity. This metric allows us to quantify the topical spread of recommendation results and to compare the diversification effects between MMR and the hybrid fine-tuning method.

Evaluation Protocol: Rank-Aware Accuracy and Beyond-Accuracy

We employed two complementary evaluation groups: Rank-aware accuracy metrics NDCG@10, MRR@10, Precision@10, and Recall@10 to assess the quality of ranked outputs.

Beyond-accuracy metrics novelty pop, intra-list diversity, and catalog coverage to capture the exploratory and coverage-oriented value of recommendations.

This dual framework follows the rank-aware novelty/diversity perspective proposed by Vargas and Castells (Vargas & Castells, 2011) and the methodological guidance outlined in recent surveys (Kaminskas & Bridge, 2016).

Each metric was computed on both validation and test splits to ensure the observed patterns generalize beyond hyperparameter optimization.

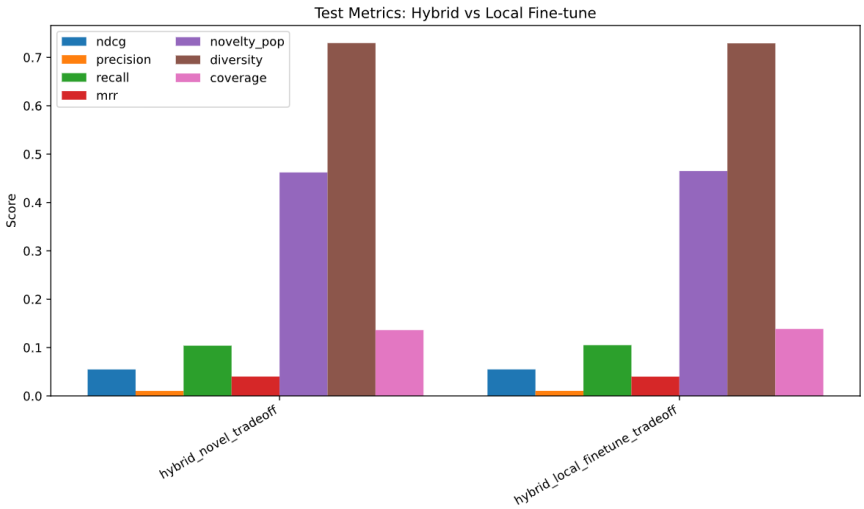

Figure 5. Comparison of evaluation metrics on test set between hybrid models and local fine-tuning approaches. The chart displays both rank-aware metrics (NDCG, Precision, Recall, MRR) and beyond-accuracy metrics (novelty, diversity, coverage) to validate the model's ability to preserve accuracy while enhancing novelty and diversity.

From Figure 5, we observe that hybrid and MMR variants demonstrate distinct behavior: the hybrid maintains near-identical accuracy while moderately enhancing novelty and coverage, whereas MMR produces a stronger diversity gain but with measurable accuracy loss. This complementary behavior supports the interpretation that local fine-tuning and classical re-ranking address the same trade-off through different mechanisms.

Hyperparameter Selection and Validation Strategy

For the hybrid model, signal weights were tuned locally around a near-optimal configuration that preserved stable accuracy while yielding measurable gains in novelty and coverage. For MMR, λ was explored across a fixed grid to capture multiple balance points. We retained two representative configurations for each method: one maximizing validation NDCG and another representing the best accuracy-novelty balance. All models were then evaluated on the held-out test split to ensure robustness of results.

Reproducibility and Experimental Artifacts

To ensure transparency and reproducibility, all experiments were conducted within a controlled and fully documented framework. The implementation includes independent modules for baseline modeling, hybrid fine-tuning, and MMR-based re-ranking, each executed under consistent experimental conditions and dataset splits. All parameter settings, evaluation scripts, and result logs were systematically versioned and validated to guarantee consistency across repeated runs. A comprehensive record of evaluation outcomes covering both rank-aware accuracy and beyond-accuracy metrics is maintained for all experimental configurations. These artifacts collectively form a reproducible pipeline that allows independent verification of our findings and facilitates future extensions or comparative studies.

As illustrated in Figure 4 on Evaluation Protocol: Rank-Aware Accuracy and Beyond-Accuracy, the comparison between hybrid variants on the test split shows that local fine-tuning effectively maintains ranking accuracy (NDCG, Precision, Recall, and MRR) while producing notable improvements in novelty and catalog coverage. Both configurations achieve nearly identical rank-aware scores, confirming that local fine-tuning does not compromise ranking quality.

Meanwhile, the moderate increases in novelty pop and coverage, together with consistently high diversity values, indicate enhanced discovery capability without degradation in core performance metrics. This balanced outcome demonstrates the stability and practical robustness of the proposed hybrid fine-tuning approach for real-world recommendation environments.

Findings and Discussion

Accuracy versus Novelty and Diversity

The experimental findings demonstrate that the hybrid recommender with local fine-tuning successfully preserves ranking accuracy while improving novelty and catalog coverage.

Across both validation and test splits, the difference in NDCG@10 remained less than 0.2% in absolute terms, indicating virtually unchanged ranking quality. At the same time, catalog coverage increased by approximately 1.7% on the test split, accompanied by a modest gain in novelty based on popularity-weighted measures. As shown in Figure 2 (introduced in Section 1) illustrates the before

and after comparison on the validation split indicates that the hybrid variants maintain stable ranking accuracy across all metrics.

The NDCG, Precision, Recall, and MRR scores remain nearly identical before and after fine-tuning, confirming that the hybrid and locally fine-tuned configurations preserve accuracy while enabling further novelty-oriented adjustments. The hybrid's improvements become more evident in discovery-oriented metrics.

As shown in Figure 2, which presents validation results for novelty and diversity, both novelty and coverage increase without any measurable degradation in rank-aware accuracy metrics. This demonstrates that incorporating user-specific novelty signals and lightweight content-based factors, combined with restrained local fine-tuning, enables the system to promote exploration while maintaining user relevance. Such balanced behavior is crucial for practical deployment, where excessive optimization may compromise model interpretability and long-term stability.

Classical Diversification using MMR

To benchmark against traditional diversification approaches, we applied Maximal Marginal Relevance (MMR) re-ranking to the ItemKNN baseline. The experimental results demonstrate that MMR successfully increases intra-list diversity by approximately 2-3 percentage points on the validation split while maintaining well-controlled accuracy performance, with only a marginal decline in NDCG@10. This controlled trade-off behavior aligns precisely with the theoretical formulation of MMR, where relevance scores are systematically balanced against similarity penalties to previously selected items, thereby effectively reducing redundancy in the ranked recommendation list.

As illustrated in Figure 2 on Introduction, the MMR validation results reveal distinct parameter sensitivity patterns across both accuracy and diversity dimensions. The left panel of Figure 2 demonstrates that higher λ values ($\lambda = 0.8$) consistently favor relevance preservation, achieving maximum NDCG performance, while the right panel shows that moderate λ values ($\lambda = 0.6$) yield the most balanced accuracy-diversity trade-off. Specifically, configurations around $\lambda = 0.8$ tend to maximize rank-aware metrics such as NDCG and precision, whereas $\lambda \approx 0.6$ achieves a more desirable equilibrium between accuracy retention and novelty enhancement, as evidenced by improved diversity and coverage metrics without substantial accuracy degradation.

These empirical observations strongly reinforce the classical theoretical understanding proposed by Carbonell and Goldstein (Carbonell & Goldstein, 1998), confirming that MMR's λ parameter serves as a direct control mechanism governing the fundamental tension between relevance optimization and diversity promotion in recommendation systems.

Rank-Aware Evaluation Framework

The use of rank-aware metrics including NDCG@10, MRR@10, Precision@10, and Recall@10 combined with beyond-accuracy indicators such as novelty_pop, intra-list diversity, and catalog coverage, provides a holistic view of recommendation utility. Following the framework of Vargas and Castells (Castells et al., 2022), these complementary measures capture both top-ranked relevance and exploratory potential within the list. Unlike accuracy-only assessments, this multidimensional perspective reveals how models perform not just in retrieving relevant items, but also in promoting item discovery and catalog reach.

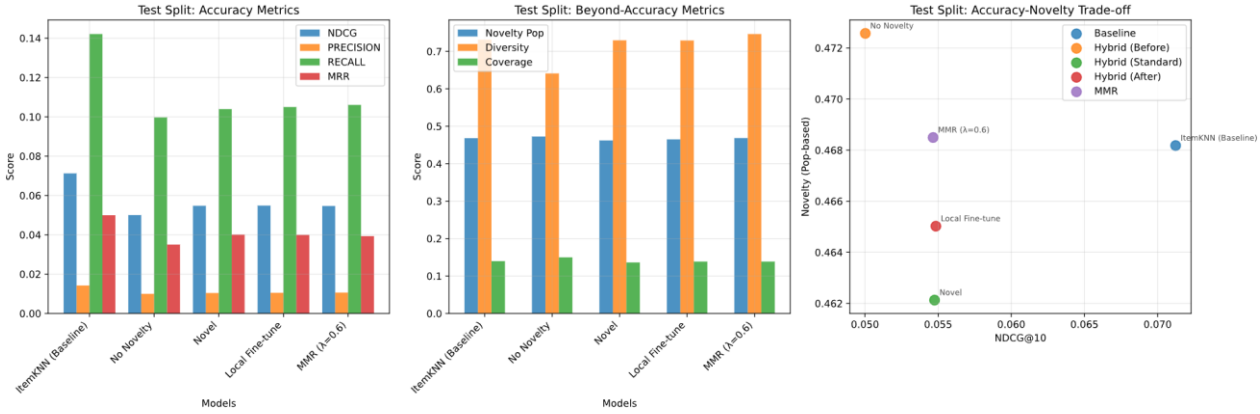

Figure 5. Test split comparisons demonstrate complementary effects of hybrid and MMR approaches on accuracy-diversity trade-offs. Panel A shows rank-aware metrics preservation across different methodologies. Panel B reveals distinct beyond-accuracy improvements: hybrid models achieve moderate novelty gains while maintaining coverage, whereas MMR delivers stronger intra-list diversification. Panel C visualizes the accuracy-novelty trade-off space, confirming that hybrid approaches preserve relevance while MMR achieves superior diversity at predictable accuracy costs, validating rank-aware evaluation principles from Beyond-Accuracy literature.

Figure 5 presents a comprehensive test split comparison that demonstrates the distinct yet complementary effects of hybrid and MMR approaches across three analytical dimensions. Panel A illustrates the preservation of rank-aware metrics, showing that both hybrid variants and MMR maintain competitive accuracy performance relative to the ItemKNN baseline. Panel B reveals the differential impact on beyond-accuracy metrics, where hybrid models achieve moderate improvements in novelty and coverage while preserving diversity, whereas MMR delivers substantially stronger intra-list diversification with enhanced novelty scores. Panel C visualizes the accuracy-novelty trade-off space, positioning each approach within the fundamental tension between relevance optimization and exploration promotion.

The comparative test results validate these multidimensional observations through quantitative evidence. The hybrid approach, particularly with local fine-tuning, preserves accuracy while moderately increasing novelty and coverage, demonstrating controlled enhancement of exploratory potential without sacrificing ranking precision. In contrast, MMR achieves stronger intra-list diversification and superior novelty scores, but at a small, predictable cost to ranking precision a

trade-off that aligns precisely with its theoretical formulation. These findings illustrate how different methodological choices shape the balance between user satisfaction and exploration, reflecting the rank-aware evaluation principles highlighted in the Beyond-Accuracy literature (Raza et al., 2022).

The positioning of models in the accuracy-novelty space (Panel C) particularly underscores the strategic value of each approach: hybrid methods occupy the high-accuracy, moderate-novelty quadrant, making them suitable for applications prioritizing relevance with controlled exploration, while MMR positions itself in the moderate-accuracy, high-novelty region, ideal for scenarios demanding maximum diversity and serendipitous discovery.

Interpretation of Key Findings

The key insights derived from the experiments can be summarized as follows:

1. Hybrid with Local Fine-Tuning: Incorporating a user-specific novelty component and combining lightweight signals (content similarity, popularity, and SVD factors) enables measurable gains in novelty and coverage without sacrificing ranking accuracy.
   The approach remains stable, interpretable, and well-suited for integration into real-world recommendation pipelines.
2. MMR over ItemKNN: Provides a straightforward, model-agnostic mechanism for controlling diversity via the λ parameter.
   As a post-ranking module, MMR requires no retraining, making it a practical option for production systems seeking diversity enhancement with minimal engineering overhead.
3. Complementarity of Approaches: The hybrid and MMR methods are not substitutes but complements:
   The hybrid enhances novelty and coverage while preserving accuracy, whereas MMR explicitly promotes intra-list diversity with an interpretable trade-off.
   Combined, they offer a flexible toolkit for balancing relevance, novelty, and diversity depending on application goals.

Operational Insights and Trade-off Analysis

From an operational standpoint, experimental results provide clear, actionable guidelines for balancing accuracy, novelty, and diversity in practical recommender-system deployment.

1. Parameter Control
   The λ parameter in Maximal Marginal Relevance (MMR) determines the balance between relevance and diversity.
   A higher λ is suitable for relevance-driven contexts where ranking precision is critical, whereas a moderate λ (≈ 0.6) yields a balanced diversity level with only a marginal drop in NDCG.

In contrast, the hybrid configuration should be fine-tuned locally around its accuracy-optimal region to maintain stability while incrementally improving novelty and catalog coverage.

2. Metric Reporting

   Comprehensive evaluation should report both rank-aware (NDCG@10, MRR@10, Precision@10, Recall@10) and beyond-accuracy metrics (novelty, diversity, coverage). Such transparent reporting aligns system performance with user-experience objectives, helping practitioners distinguish between improvements in ranking accuracy and genuine gains in user discovery and engagement.

3. Comparative Results

   Table 1 summarizes the test-split outcomes, contrasting hybrid and MMR approaches. Hybrid variants maintain near-perfect ranking accuracy while slightly enhancing novelty and coverage.

   Conversely, MMR provides superior diversity through controlled trade-offs in accuracy, giving developers distinct operational choices between stability-focused and diversity-focused strategies.

Table 1. Test split performance summary showing complementary strengths of hybrid and MMR strategies. Hybrid approaches maintain near-perfect ranking performance while incrementally enhancing novelty, whereas MMR achieves superior diversity metrics at predictable accuracy costs, providing practitioners with distinct operational choices.

Model	NDCG@10	Precision@10	Recall@10	Novelty	Diversity	Coverage	Strategy
ItemKNN (Baseline)	0.0712	0.0142	0.1421	0.4682	0.7313	0.1397	Baseline
Hybrid (Before)	0.0500	0.0100	0.0997	0.4726	0.6411	0.1498	Stability-focused
Hybrid (Standard)	0.0548	0.0104	0.1039	0.4621	0.7296	0.1361	Standard
Hybrid (After)	0.0548	0.0105	0.1050	0.4650	0.7291	0.1385	Balanced
MMR (λ = 0.6)	0.0547	0.0106	0.1060	0.4685	0.7464	0.1385	Diversity-focused
Note: *Higher values indicate better performance for all metrics. Hybrid approaches occupy the stability-focused region, while MMR extends the diversity frontier.*							

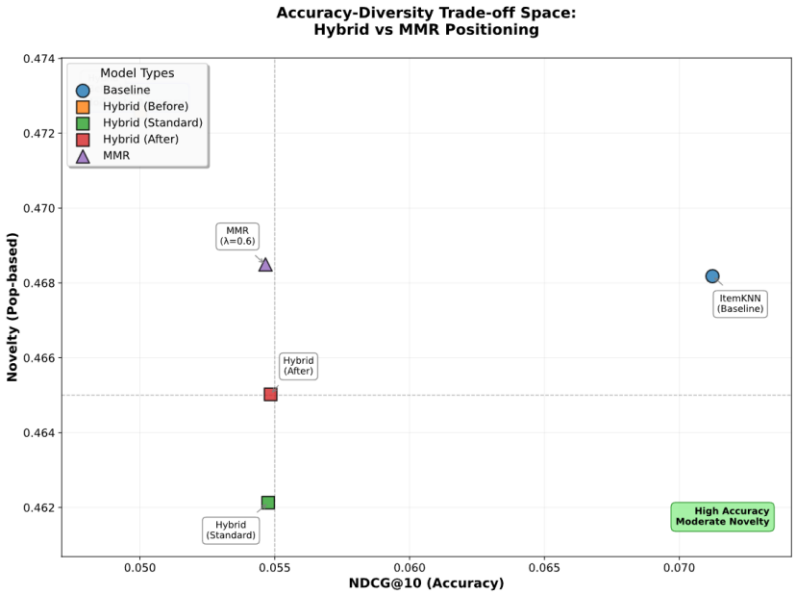

Figure 6. Accuracy-diversity trade-off space demonstrating strategic positioning of hybrid and MMR approaches. The hybrid model occupies the stability-focused corner (high accuracy, moderate novelty), while MMR extends the diversity frontier (moderate accuracy, high novelty), jointly mapping the feasible region between accuracy and exploration.

As shown in Figure 6, which plots the accuracy-novelty trade-off space, the hybrid models cluster closely around the stability region, maintaining accuracy with moderate novelty gains.

The MMR ($\lambda = 0.6$) configuration shifts upward and slightly leftward, illustrating its ability to enhance diversity and novelty with predictable accuracy reduction.

Together, these strategies span the feasible region between accuracy preservation and exploration capability, giving practitioners a tunable framework for operational decision-making.

Validity and Limitations

While the results presented are consistent and statistically stable, several limitations warrant consideration. First, the findings are based on a single benchmark dataset, and outcomes may vary in domains with different popularity or genre dynamics. Second, both the hybrid weights and MMR's λ parameter influence the shape of the accuracy-novelty trade-off; thus, their sensitivity should be examined under varying data distributions. Future work should extend the evaluation to multi-domain datasets and investigate adaptive mechanisms for dynamically tuning these parameters in response to user feedback and catalog evolution.

Conclusion and Suggestion

This study shows that ranking accuracy can be preserved while improving novelty and catalog coverage through a hybrid recommender with local fine-tuning. Across validation and test sets, NDCG@10 differed by less than 0.2%, while coverage increased by approximately 1.7% and novelty rose moderately. Thus, the hybrid effectively enhances discovery without compromising accuracy.

The MMR re-ranking experiment further confirms that diversification can be precisely controlled. By adjusting λ, MMR increased intra-list diversity by 2-3 percentage points with a small, predictable accuracy loss consistent with classical relevance-diversity trade-offs. Key contributions include:

a) a simple, interpretable hybrid design with user-specific novelty;
b) local fine-tuning near the optimal accuracy-novelty balance to prevent overfitting; and
c) a rank-aware and beyond-accuracy evaluation framework integrating novelty, diversity, and coverage.

Practically, the hybrid approach suits real-world deployment, maintaining accuracy while improving exploration value, whereas MMR can serve as a lightweight post-ranking module to adjust diversity without retraining. Tuning hybrid weights or λ provides flexible control aligned with user and business goals. Future work should extend evaluation across domains, include user-centric and fairness analyses, and explore adaptive or explainable mechanisms for balancing relevance and discovery. Overall, the findings demonstrate that accuracy and novelty are not conflicting objectives when optimization emphasizes interpretability, localized tuning, and transparent evaluation.

References

Bertani, R. M., A. C. Bianchi, R., & Costa, A. H. R. (2020). Combining novelty and popularity on personalised recommendations via user profile learning. *Expert Systems with Applications*, *146*, 113149. https://doi.org/10.1016/J.ESWA.2019.113149.

Carbonell, J., & Goldstein, J. (1998). The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. *SIGIR 1998 - Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 335–336. https://doi.org/10.1145/290941.291025.

Castells, P., Hurley, N., & Vargas, S. (2022). Novelty and Diversity in Recommender Systems. *Recommender Systems Handbook: Third Edition*, 603–646. https://doi.org/10.1007/978-1-0716-2197-4_16.

Hurley, N., & Zhang, M. (2011). Novelty and Diversity in Top-N Recommendation -- Analysis and Evaluation. *ACM Transactions on Internet Technology (TOIT)*, *10*(4). https://doi.org/10.1145/1944339.1944341.

Isufi, E., Pocchiari, M., & Hanjalic, A. (2021). Accuracy-diversity trade-off in recommender systems via graph convolutions. *Information Processing & Management*, *58*(2), 102459. https://doi.org/10.1016/J.IPM.2020.102459.

Jacek Wasilewski, N. H. (n.d.). *Incorporating Diversity in a Learning to Rank Recommender System*. Retrieved October 15, 2025, from https://aaai.org/papers/572-flairs-2016-12944/.

Javari, A., & Jalili, M. (2014). Accurate and Novel Recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(4), 1–20. https://doi.org/10.1145/2668107.

Kaminskas, M., & Bridge, D. (2016). Diversity, Serendipity, Novelty, and Coverage. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*(1). https://doi.org/10.1145/2926720.

Ma, M., & Jiang, Y. (2020). A meta-level hybrid recommendation method based on user novelty. *ACM International Conference Proceeding Series*, *PartF168986*, 616–625. https://doi.org/10.1145/3452940.3453060.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: How accuracy metrics have hurt recommender systems. *Conference on Human Factors in Computing Systems - Proceedings*, 1097–1101. https://doi.org/10.1145/1125451.1125659.

Raza, S., Bashir, S. R., & Naseem, U. (2022). *Accuracy meets Diversity in a News Recommender System* (pp. 3778–3787). https://aclanthology.org/2022.coling-1.332/.

Vargas, S., & Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. *RecSys'11 - Proceedings of the 5th ACM Conference on Recommender Systems*, 109–116. https://doi.org/10.1145/2043932.2043955.

Xia, L., Xu, J., Lan, Y., Guo, J., & Cheng, X. (2015). Learning maximal marginal relevance model via directly optimizing diversity evaluation measures. *SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 113–122. https://doi.org/10.1145/2766462.2767710.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). *Improving recommendation lists through topic diversification*. 22. https://doi.org/10.1145/1060745.1060754.