Integrating Stochastic Models into News Recommender Systems: A Comparative Study with Hybrid Approaches

Bráulio M. Veloso[1], **Anderson A. Ferreira**[1]

[1] Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Ouro Preto (UFOP) – Minas Gerais – Brasil

braulio.veloso@aluno.ufop.edu.br, anderson.ferreira@ufop.edu.br

Abstract. *Recommender systems are essential for handling information overload in online newspapers, where the lack of explicit feedback and the need for real-time updates limit computationally expensive methods. This work investigates how contextualization based on stochastic models that capture the sequential and uncertain nature of reading behavior improves news recommender systems. Using real access logs and a simulation environment, we evaluate these models across multiple metrics. The results show that stochastic models outperform traditional approaches, especially in personalization and diversity, and that combining them with hybrid methods yields the best performance.*

1. Introduction

Recommender systems play a crucial role in mitigating information overload on the Web, where the continuous increase in available content makes it impractical for users to manually identify relevant information [Wu et al. 2023, Pu and Beam 2024]. On online news platforms, these systems are essential both on individual newspaper websites and on large-scale news aggregators [Bai et al. 2017, Lv et al. 2024]. News platforms operate under a constant flow of new items, frequent updates, and the removal of outdated content, requiring recommendation strategies that can adapt to this rapid turnover. Because explicit feedback is rarely available, interactions are inferred from implicit signals such as access logs, reading time, and navigation traces [Raza and Ding 2022, Xin et al. 2022].

News Recommender Systems (NRS) must operate under several domain-specific constraints. *Implicit and noisy interaction data:* User behavior is usually short, anonymous, and composed of low-interaction signals, resulting in ambiguous and noisy feedback that complicates preference inference [Epure et al. 2017, Wu et al. 2023]. *High-volume and changing content:* News platforms continually introduce new content, update existing items, and remove outdated ones [Campos et al. 2014, Karimi et al. 2018]. Relevance is highly time-sensitive, and user interest shifts quickly in response to breaking events [Cavalcante and Pinheiro 2013, Moreira et al. 2018]. *Double cold-start scenario:* new users with new items creating a double cold-start scenario [Meng et al. 2023, Pu and Beam 2024]. *Real-time:* The system must generate recommendation lists under strict latency requirements while new items are constantly added [Ren and Shi 2022]. To meet these constraints, many platforms rely on simple heuristics such as recommending the most popular or most recent items [Das et al. 2007, Symeonidis et al. 2022], which favors speed at the expense of personalization and diversity. Collaborative-filtering methods struggle under rapidly outdated preferences and sparse interaction matrices [Ji et al. 2021, Li et al. 2022], while content-based approaches have difficulty adapting to dynamic shifts in topic relevance and news importance [Li et al. 2014, Symeonidis et al. 2022].

Despite recent advancements, a key challenge remains: modeling user reading behavior as a sequential, uncertain process to improve contextualization. Naive approaches ignore session dynamics, whereas many advanced ones depend on static user profiles or computationally intensive pipelines. There is limited research on lightweight probabilistic methods capable of capturing local sequential transitions within reading sessions, operating under high item turnover and limited feedback, adapting to anonymous short-lived interactions, and providing real-time responses with low computational cost. This gap motivates investigating stochastic models to contextualize the NRS.

In this work, we examine how stochastic-model-based contextualization improves NRS performance and whether it outperforms alternative strategies. Using a real-world dataset of user access logs and a simulation environment that reconstructs reading sessions, we compare stochastic modeling approaches against standard baselines across accuracy, coverage, diversity, novelty, personalization, and serendipity metrics. The main contribution of this paper is a comprehensive study of the impact of integrating stochastic models into recommendation systems under various evaluation criteria. Our results show that stochastic models outperform baseline methods, including independence-based and category-filtering strategies. Moreover, hybrid recommenders consistently outperform direct integration of stochastic models into base recommenders, and combining stochastic models with hybrid methods yields further improvements. Notably, the most expressive gains appear in personalization and diversity, underscoring the value of probabilistic modeling in capturing meaningful behavioral patterns in news consumption.

2. Related Works

In a scenario marked by information overload and dispersed audiences, personalized recommendations offer an effective way to capture users' individualized interests, increase retention, and drive positive commercial impact [Karimi et al. 2018]. While non-personalized solutions assume homogeneous user interests and risk saturating audiences with overly explored topics, personalized systems adapt content selection to each user's profile and context, providing more relevant and engaging experiences [Liu et al. 2010, Bai et al. 2017]. Early studies attempted to adapt classical collaborative techniques to the news domain, but the lack of explicit feedback and the need for real-time responses soon revealed the limitations of such methods [Das et al. 2007]. Other works employed popularity and temporal trends to address recency effects [Garcin et al. 2013], but although effective in the short term, these approaches often lead to redundancy and bias toward sensationalist content [Cavalcante and Pinheiro 2013, Koo et al. 2021].

Unlike traditional recommender systems in e-commerce, where preferences tend to be stable, NRS must operate under highly dynamic content and rapidly shifting user interests [Bai et al. 2017, Karimi et al. 2018, Koo et al. 2021]. Strategies such as probabilistic click modeling and reading-time analysis enable more accurate estimations of user engagement, distinguishing between superficial and in-depth interactions [Constantinides and Dowell 2018, Ji et al. 2021]. Some works further focus on enhancing the interpretation of implicit feedback and enriching contextual representations to capture evolving preferences in fast-paced news environments [Li et al. 2014, Sheu et al. 2022]. More recently, some works employed deep learning models and advanced semantic representations to capture content nuances and evolving preferences, though at substantially higher computational cost [Moreira et al. 2018, Li et al. 2022, Wang et al. 2025].

Understanding readers' interactions with content helps anticipate and personalize preferences [Lv et al. 2024]. News consumption is highly volatile: users frequently seek updates on ongoing events or explore diverse topics [Symeonidis et al. 2022]. Empirical studies highlight recurring reading patterns that influence personalization [Epure et al. 2017, Park 2025]. For example, [Li et al. 2022] shows that many readers display episodic behaviors, returning to follow unfolding events, or adopt thematic exploration patterns by navigating across multiple topics. [Lv et al. 2024] investigates the interplay between geographic and semantic features, concluding that semantic aspects often exert stronger influence, as both dimensions shape thematic preferences. Additional forms of implicit feedback (such as time spent on page) provide valuable clues about user interest [Ji et al. 2021]. [Constantinides and Dowell 2018] employs probabilistic algorithms and neural networks to predict next-user actions, whereas [Symeonidis et al. 2022] leverages Random Walks on time-evolving heterogeneous networks; [Ren and Shi 2022] integrates auxiliary news information and temporal user–news interactions through heterogeneous graphs to model temporal click sequences.

Accurately modeling user behavior in news contexts remains an open challenge, particularly when balancing personalization with informational diversity and accounting for contextual factors such as time, device, or mood [Campos et al. 2014, Meng et al. 2023]. Moving beyond implicit feedback alone, [Bai et al. 2017] incorporates extrinsic data, such as browsing and search histories, to build personalized profiles. In contrast, our approach focuses on internal contextual signals derived from the platform itself. [Li et al. 2022] embeds category information in a neural model using a poly-attention mechanism to learn multiple user-interest vectors, capturing different facets of preferences. Like that work, we assume that user behavior evolves over time and that a model should adapt accordingly; our stochastic model progressively learns from user interactions, adjusting to emerging reading patterns. [Epure et al. 2017] examines short-, medium-, and long-term interest shifts, showing that combining multiple temporal windows improves accuracy and diversity. However, their modeling relies solely on a first-order Markov process and remains limited to offline behavioral analysis, without demonstrating integration into an operational NRS. [Veloso et al. 2019] investigates several stochastic models to identify those that best capture transitions between reading categories; we adopt some of these models to represent user behavioral dynamics throughout reading sessions in a recommendation environment.

3. Methodology

We developed a Python simulation framework[1] to reproduce user interactions with a digital news platform in chronological order. We evaluated existing frameworks, but none of them support the combination of progressive user arrivals, dynamic pool updates, session-aware modeling, and integration with stochastic models. Our framework ensures that each recommendation reflects the actual set of news available at the time of access and progressively introduces newer users with individualized reading behaviors. At each user access, the system uses the user's history and the current news pool to generate a recommendation list, and we compare it with the next item in the same session that the user actually reads. This iterative process enables detailed metric computation and systematic comparison. A

[1] available at `https://github.com/bmveloso/CNRS-SM`

data preparation step splits the dataset into training and test subsets. The framework processes the training set to initialize both the news recommenders and the stochastic models, then partitions the test set into P chronological folds. Within each interval, chronologically ordered accesses generate recommendation requests, and the corresponding metrics evaluate NRS performance. Once the framework processes a fold, it incorporates that fold's data into the models and the systems before moving to the next one.

We implemented a set of NRS that cover non-personalized and personalized methods, in both pure and hybrid variants, selected for their simplicity, interpretability, and wide adoption. Our objective is not to develop new algorithms but to evaluate how stochastic models affect different recommendation strategies. Traditional NRS provides a controlled and meaningful baseline for integration. **Non-personalized** approaches include *Popularity filtering* (`Pop`), which recommends the most accessed news, and *Recency filtering* (`Rec`), which selects the most recently published news at that time. **Personalized** approaches comprise *Co-occurrence filtering* (`Coo`), which counts how often pairs of news items appear consecutively within user sessions and recommends those with the highest co-occurrence strength with the target item, and *Item-based collaborative filtering* (`ICF`), which constructs an item-to–item cosine similarity matrix considering only items with at least five readers to avoid sparsity and to control size. We applied two hybridization strategies following [Çano and Morisio 2017]. (i) Feature-based hybridization for non-personalized systems: `Pop_H` periodically readjusts popularity scores using a time-decay factor to emphasize recent interactions, and `Rec_H` augments recency-based ranking with an additional popularity feature, selecting the most popular items within a recency window. (ii) Complement-based hybridization for personalized systems, where `Pop` is used to fill incomplete lists (`Coo_H`, `ICF_H`).

To study how sequential patterns in reading behavior enhance recommendations, we incorporated Stochastic Models (SM). Following prior work [Veloso et al. 2019], we selected five models from two families, Higher-Order Markov Models and Past-Function Models. In all cases, news categories define the states, and models are trained on past accesses to predict probabilities during each session.

Higher-Order Markov Models assume that the next category in a reading sequence depends on a limited number of recent observations. This class of models captures local temporal dependencies. While lower-order models condition predictions on only one or two past categories, this study focuses on *Third-Order* (`Mk3`) and *Fourth-Order* (`Mk4`) models, which provide a compact but rich contextual window. Formally, given a sequence of visited categories $\{C_1, C_2, \ldots, C_i\}$, an n-order Markov model estimates the probability of the next category C_{i+1} as:

$$\mathbb{P}(C_{i+1} \mid C_1, C_2, \ldots, C_i) = \mathbb{P}(C_{i+1} \mid C_{i-(n-1)}, C_{i-(n-2)}, \ldots, C_i). \quad (1)$$

Past-Function Models consider the user's entire interaction history, capturing indirect and non-sequential behavioral patterns by applying heuristic functions that estimate the probability of the next category. Let $\{l_1, \ldots, l_m\}$ denote all possible category labels, and $\mathbf{f}(\bullet)$ be the prediction function for the next category. Under this formulation, a past-function model estimates the probability of the next category as:

$$\mathbb{P}(C_{i+1} \mid C_1, ..., C_i) = \mathbb{P}(C_{i+1} \mid \mathbf{f}(l_1) = x_1, ..., \mathbf{f}(l_m) = x_m) \propto \mathbb{P}(C_{i+1} \mid \mathbf{f}(\bullet) = x) \quad (2)$$

The proportionality symbol ($\propto$) indicates that values are normalized so probabilities sum to 1. We implemented three variants: (1) The *Number of Visits Model* (`NV`) evaluates how frequently each category $l \in \{l_1, \ldots, l_m\}$ has appeared in the session, using the cumulative count as a proxy for user interest, with $\mathbf{f}(\bullet) = S_i^l = \sum_{j=1}^{i} \mathbf{I}[C_j = l]$ and $x \in [0, i]$. (2) The *Readings After Departure Model* (`RAD`) measures how many items the user reads after leaving a category l, defined by $\mathbf{f}(\bullet) = Q_i^l$, where $Q_i^l = -1$ if l is not visited, $Q_i^l = 0$ if the current item belongs to category l, and otherwise $Q_i^l = i - \max\{j \in [1, i] : C_j = l\}$, taking values for $x \in [-1, 0, \ldots, i-1]$. (3) And, the *Current Visit Duration Model* (`CVD`) captures the length of the user's most recent consecutive reading streak in category l. Let V_i^l be the duration of the current visit to l, where $V_i^l = 0$ if l has never been read, $V_i^l = i$ if the session has remained in l since the beginning, and otherwise $V_i^l = i - \max\{j \in [1, i] : C_j \neq l\}$. Them $\mathbf{f}(\bullet) = V_i^l$, with $x \in [0, i]$.

We also compare our main proposals for integrating stochastic models with NRS against two alternatives. The first one, *Category Filtering model* (`CatF`), assigns probability 1 to the next category being the same as the current one and 0 to the others. The second, *Independence model* (`Ind`), assigns to each category a probability proportional to its empirical reading frequency over the historical record, ignoring the user's current session or any sequential structure. These approaches serve as contrastive references, allowing us to better interpret the range and limits of the metrics.

For all models, we used the probability distribution generated by the SM to sample k categories. For each recommendation request, the SM samples categories, and the NRS selects the best item from each sampled category. Due to the probabilistic nature of the sampling, category frequencies converge to the SM's distribution over time. To integrate SM category-level probabilities with NRS item-level rankings, we applied pre- and post-filtering strategies [Champiri et al. 2015]. In pre-filtering, the SM selects some categories, and dedicated recommender systems generate recommendations proportionally to those selected categories. In post-filtering, a single recommender system ranks all items, which are then filtered based on the selected categories inferred from the stochastic models. We illustrate these two modes in Figure 1. For NRS capable of category-specific rankings (e.g., `Pop`, `Rec`, `Pop_H`, `Rec_H`), we applied *pre-filtering*, maintaining separate model instances per category and activating only the sampled ones. For systems that produce global rankings (e.g., `Coo`, `ICF`, `Coo_H`, `ICF_H`), we used *post-filtering*, in which the system filters the global output to retain only the items belonging to the categories selected by the SM. The NRS without and with SM generate a recommender list of the same size.

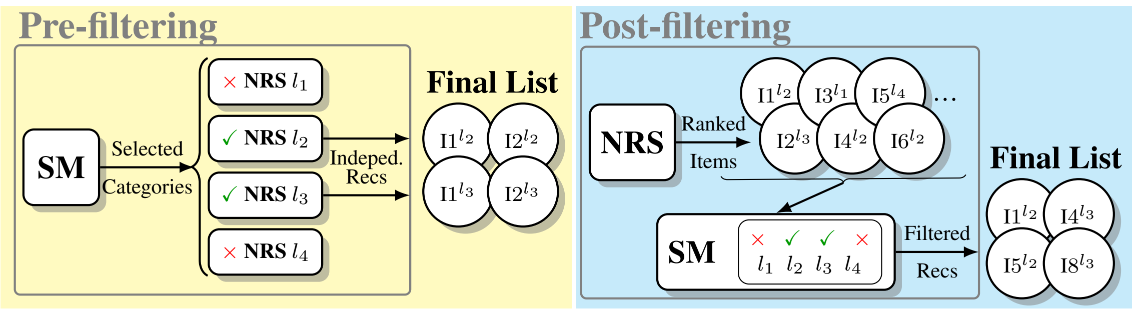

Figure 1. Stochastic models integration: Pre- and Post-filtering. This example considers four categories $\{l_1, l_2, l_3, l_4\}$ and recommends four items ($k = 4$).

4. Experiments

This study uses real data from a Brazilian online newspaper (name withheld for confidentiality), publicly available at `https://forms.gle/kvqoa78jVWEHTcEx5`. The dataset contains over 26.7 million accesses by 7.2M users to 270K reading news items across 14.4M sessions in 14 days; users spend 184.4 seconds per item and read two to three items per session on average. The news spans 10 imbalanced categories; see Table 1. For evaluation, we split the data chronologically: the first week served as training data (4.4M users, 14.3M accesses), and the second week as test data (3.8M users, 12.4M accesses). Due to the platform's streaming nature, both partitions share users and items, and the item catalog grows over time from 362k to 370k items. The test dataset was divided into $P = 28$ folds of consecutive 6-hour windows, preserving the natural interaction flow. These windows match the late night (00:00 am to 05:59 am), morning (06:00 am to 11:59 am), afternoon (00:00 pm to 05:59 pm), and evening (06:00 pm to 11:59 pm). This partition also reflects the distinct temporal access patterns observed in news consumption, where traffic is typically lowest during the late-night hours, increases in the morning, peaks in the afternoon, and declines again in the evening. All experiments used a recommendation list size of $k = 4$ and 10 news categories, following the real newspaper.

Table 1. Category-wise percentages across the entire dataset.

Category	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}
Publication (%)	0.9	3.2	12.4	1.6	32.0	3.8	6.1	30.2	6.2	3.6
Reading (%)	0.1	13.7	0.9	0.8	12.2	41.4	2.7	16.7	7.4	4.1
Recommendation (%)	5.4	6.5	0.9	0.4	11.9	45.2	5.5	15.7	8.4	0.1
Click(%)	0.001	6.4	0.3	0.2	11.1	55.5	5.0	15.0	6.4	0.04

We evaluated the NRS performance using a comprehensive set of metrics. Next, we present the metrics [Maksai et al. 2015, Garcin et al. 2013, Raza and Ding 2022] we implemented in our framework, along with their formal definitions. Notation: U: set of users; I: set of items (news); A: set of accesses as tuples (u, i, t), user u, item i and time t; $R_{u,t}$: recommendation list for user u at time t; $I_{u,t}$: item read by user u at time t; $H_{u,t}$: set of items read by user u before the time t; S: sessions (u, r, h), with r recommended items and h items read during the session; P_u^r, P_u^h: category distributions of recommended and historical items; $d(i, j)$: dissimilarity between items i and j; $D(\cdot, \cdot)$: distribution distance; $pop(i, t)$: popularity of item i at time t.

- **Hit Rate of Users** (`HRU`) Proportion of users who click at least once on the recommended items: HRU $= \frac{1}{|U|} \sum_{u \in U} \mathbf{I}\left\{\exists t \in A | I_{u,t+1} \in R_{u,t}\right\}$.

- **Catalog Coverage** (`CC`) Fraction of the catalog recommended at least once: CC $= \frac{1}{|I|} |\bigcup_{(u,i,t) \in A} \{i \in R_{u,t}\}|$.

- **Intra-List Diversity** (`ILD`) Average pairwise dissimilarity within recommendation lists. We used the Jaccard dissimilarity as $d(i, j)$ on the sets of terms extracted from news titles, as it provides a natural and interpretable measure of content diversity. ILD $= \frac{1}{|A|} \sum_{(u,i,t) \in A} \frac{1}{|R_{u,t}|(|R_{u,t}|-1)} \sum_{\substack{i,j \in R_{u,t} \\ i \neq j}} d(i, j)$.

- **Extra-List Diversity** (`ELD`) Dissimilarity between consecutive recommendation lists: ELD $= \frac{1}{|A|} \sum_{(u,i,t) \in A} \left(1 - \frac{|R_{u,t} \cap R_{u,t+1}|}{|R_{u,t} \cup R_{u,t+1}|}\right)$.

• **Expected Profile Distance** (`EPD`) Difference between the recommendation and historical category profiles of each user: EPD $= \frac{1}{|S|}\sum_{(u,r,h)\in S} D\big(P_u^r,\, P_u^h\big)$.

• **Personalized** (`Pers`) Inter-user dissimilarity of recommendation lists. We use a sliding window of 5 users to avoid the high cost of exhaustive pairwise comparisons, thereby reducing computational overhead while maintaining a representative estimate of personalization. Pers $= \frac{1}{\sum_{u\in U}|U_{\text{sampled}}(u)|}\sum_{u\in U}\sum_{v\in U_{\text{sampled}}(u)}\left(1-\frac{|R_u\cap R_v|}{|R_u\cup R_v|}\right)$.

• **Average Unpopularity Score** (`AUS`) Measures the system's tendency to recommend less popular items: AUS $= \frac{1}{|A|}\sum_{(u,i,t)\in A}\left[1-\frac{1}{|R_{u,t}|}\sum_{j\in R_{u,t}} pop(j,t)\right]$.

• **Unexpectedness** (`Unexp`) Frequency of recommendations that are both unseen and unpopular: Unexp $= \frac{1}{|A|}\sum_{(u,i,t)\in A}\frac{1}{|R_{u,t}|}\sum_{j\in R_{u,t}} \mathbf{I}\{j\notin H_{u,t}\}\,(1-pop(j,t))$.

• **Percentage of Complete List** (`PCL`) Frequency with which the system outputs full-length lists: PCL $= \frac{1}{|A|}\sum_{(u,i,t)\in A}\mathbf{I}\{\,|R_{u,t}|=k\,\}$. The cold start ratio is $1-PCL$.

We measured three time-based metrics: *accumulation* (`AccT`), *recommendation* (`RecT`), and *update time* (`UpdT`). We run all experiments on a CentOS Linux 7 server equipped with two Intel Xeon E5620 CPUs (16 threads, 2.4 GHz) and 110GB of RAM.

5. Results

We group the evaluation metrics into thematic sets, called dimensions. Tables 2, 3, 4, and 5 present the results for the Accuracy-Coverage (`HRU`, `CC`), Diversity (`ILD`, `ELD`), Novelty-Personalization (`EPD`, `Pers`), Discovery (`AUS`, `Unexp`, `PCL`) and Time (`AccT`, `RecT`, `UpdT`) dimensions for pure and hybrid systems. All reported values are averages with their standard deviations across 28 temporal folds. Except for the Time metrics, which are in seconds, all metrics fall in the range $[0, 1]$ (higher is better) and appear as percentages with 3 decimal places. We compare each contextualized system (NRS+SM) against its corresponding baseline (SRN): triangles indicate whether the contextualized system improves (▲) or decreases (▼) relative to the baseline; filled triangles (▲, ▼) denote statistically significant differences (Wilcoxon signed-rank test, $p < 0.05$), while hollow ones (△, ▽) indicate non-significant changes. We **bold** the best value achieved per NRS. Additionally, we highlight results on the Pareto frontier for each dimension and NRS+SM configuration, representing the non-dominated trade-offs with a gray background.

Across all base systems, the incorporation of SM produces consistent structural shifts in performance: `Pop` experiences the most substantial transformation, with strong gains in `HRU`, `CC`, `ILD`, and `ELD`, alongside marked reductions in `EPD` and sharp increases in `Pers`, improving discovery (`AUS`, `Unexp`) without harming `PCL`. In `Rec`, SM raises `HRU` and diversity, especially `ELD`, while slightly reducing `CC`, moderating extreme novelty, and increasing personalization with minimal time overhead. For `Coo`, `HRU`, and `CC` decline, but both diversity metrics rise substantially, `Pers` consistently improves, and discovery shifts toward higher `AUS` but lower `Unexp` and `PCL`, indicating greater personalization at the cost of accuracy and cold-start robustness. In `ICF`, SM reduces `HRU` but increases `CC`, moderately enhances `ILD` and `ELD`, and improves novelty-personalization (`EPD`, `Pers`) and discovery (`AUS`, `Unexp`), accompanied by a slight rise in cold start and stable time costs.

Comparing our five stochastic models (`Mk3`, `Mk4`, `NV`, `RAD`, `CVD`) with the two ad-

Table 2. Results of pure news recommender systems in isolation and integrated with stochastic models across dimensions, part 1.

		Accuracy-Coverage (%)		*Diversity (%)*		*Novelty-Personalization (%)*	
NRS	**SM**	**HRU**	**CC**	**ILD**	**ELD**	**EPD**	**Pers**
Pop		10.567±11.90	0.001±46e-7	60.741±22.58	3.334±3.82	**52.454±19.44**	5.163±6.01
Pop	CatF	▲**16.415±12.19**	▲**0.014±60e-7**	▽45.136±3.13	▲64.417±11.57	▼12.391±7.13	▲61.725±12.02
Pop	Indep	▽10.468±9.56	▲0.010±81e-7	▲**81.138±1.70**	▲62.663±2.32	▼26.027±8.84	▲57.173±1.12
Pop	Mk3	▲15.083±11.68	▲**0.014±79e-7**	▲71.064±4.07	▲67.264±6.79	▼17.729±7.46	▲62.332±4.91
Pop	Mk4	▲15.102±11.68	▲0.013±80e-7	▲71.081±4.20	▲**67.609±6.94**	▼17.437±7.36	▲62.428±4.96
Pop	NV	▲14.522±11.30	▲0.013±98e-7	▲72.718±3.92	▲66.817±6.46	▼17.558±7.48	▲62.534±4.13
Pop	RAD	▲14.446±11.24	▲**0.014±95e-7**	▲72.707±3.50	▲67.103±5.84	▼18.101±7.66	▲**62.664±3.95**
Pop	CVD	▲13.977±11.05	▲0.013±96e-7	▲74.698±3.00	▲66.811±4.67	▼19.637±7.50	▲62.200±3.71
Rec		0.020±0.02	**0.121±0.09**	53.282±11.92	0.173±0.22	**76.804±15.26**	69.880±16.86
Rec	CatF	▲**0.587±1.21**	▼0.072±0.05	▼31.659±4.73	▲61.102±13.51	▼65.378±19.49	△72.148±11.63
Rec	Indep	▲0.091±0.20	▼0.088±0.05	▲**65.573±4.05**	▲61.556±2.18	▼66.286±11.26	△74.435±6.64
Rec	Mk3	▲0.343±0.83	▼0.097±0.06	△56.579±4.54	▲65.147±7.92	▼65.990±15.72	△74.959±7.54
Rec	Mk4	▲0.340±0.83	▼0.097±0.06	△56.637±4.61	▲**65.499±8.09**	▼65.759±15.87	△74.999±7.55
Rec	NV	▲0.336±0.86	▼0.094±0.06	△58.074±4.34	▲64.823±7.48	▼65.486±15.68	▲75.446±7.04
Rec	RAD	▲0.340±0.84	▼0.094±0.06	△58.028±4.22	▲65.104±6.83	▼65.716±15.14	▲**75.483±6.94**
Rec	CVD	▲0.322±0.79	▼0.095±0.06	▲59.827±4.09	▲64.948±5.44	▼65.728±14.24	▲75.478±6.85
Coo		**64.958±6.08**	**6.910±2.16**	64.405±5.64	88.864±7.02	10.832±5.29	80.499±8.88
Coo	CatF	▼55.558±5.63	▼4.698±1.30	▼44.228±2.10	▲91.638±5.72	▼8.753±5.43	▲85.289±7.81
Coo	Indep	▼48.485±6.68	▼5.003±1.38	▲**81.041±1.40**	▲92.439±3.35	▲**19.043±8.30**	▲86.093±3.69
Coo	Mk3	▼57.394±5.36	▼5.637±1.60	▲69.663±3.73	▲93.096±4.00	▽10.655±5.04	▲87.022±5.05
Coo	Mk4	▼57.372±5.37	▼5.652±1.62	▲69.579±3.85	▲93.131±3.99	▽10.430±4.98	▲87.035±5.06
Coo	NV	▼56.778±5.42	▼5.671±1.65	▲71.552±3.56	▲93.033±3.93	▽10.316±5.13	▲87.263±4.64
Coo	RAD	▼56.847±5.38	▼5.652±1.61	▲71.325±3.12	▲93.182±3.83	▽11.079±5.42	▲**87.310±4.62**
Coo	CVD	▼56.362±5.48	▼5.423±1.46	▲73.272±2.72	▲**93.240±3.62**	△12.630±5.63	▲87.258±4.55
ICF		**50.663±6.94**	3.441±0.51	60.575±4.42	91.681±5.76	10.226±5.80	85.034±7.51
ICF	CatF	▼40.726±5.61	▲3.809±0.59	▼44.164±2.08	▽91.459±6.00	▽9.049±5.01	▽84.918±7.87
ICF	Indep	▼28.736±6.33	▲3.593±0.65	▲**79.467±1.71**	▲**93.998±3.37**	▲**18.608±7.76**	▲87.440±4.85
ICF	Mk3	▼38.154±6.33	▲3.957±0.68	▲67.796±3.80	▲93.491±4.36	△11.002±5.02	▲87.203±5.94
ICF	Mk4	▼38.129±6.35	▲3.946±0.67	▲67.618±3.88	▲93.510±4.36	△10.814±4.97	▲87.202±5.95
ICF	NV	▼36.927±6.32	▲3.906±0.65	▲69.759±3.78	▲93.637±4.21	△10.762±4.96	▲**87.539±5.59**
ICF	RAD	▼37.141±6.28	▲**3.975±0.66**	▲69.502±3.38	▲93.658±4.15	△11.385±5.19	▲87.533±5.58
ICF	CVD	▼36.614±6.30	▲3.950±0.71	▲71.197±3.04	▲93.782±3.95	▲12.589±5.37	▲87.493±5.55

versarial models (`Indep`, `CatF`) across dimensions, we observe clear distinctions in the Pareto frontiers. In the accuracy–coverage dimension, our models remain tied with `CatF`, presenting the same number of frontier cases. In the diversity and novelty–personalization dimensions, our models systematically coincide with `Indep` on the frontiers, while consistently outperforming `CatF`, which does not reach frontier positions for these metrics. In the discovery dimension, the relative position shifts according to the base recommender: for `Pop`, `CatF` dominates the frontier; for `Rec`, the advantage moves to `Indep`; and in the two personalized NRS, our models share the frontier, sometimes aligned with `CatF`, sometimes with `Indep`, reflecting a greater dependence on the base system in this dimension. Overall, our stochastic models exhibit frontier-level performance in most dimensions, particularly in those where diversity and personalization play a central role.

Comparing the best-performing contextualized pure NRS configurations enhanced with SM (Tables 2 and 3) to their corresponding NRS hybrid versions, without SM (Tables 4 and 5), several distinctions emerge. In `HRU`, hybridization exerts a stronger impact, yielding substantially higher gains than stochastic contextualization alone. For `CC`, however, the use of SM surpasses hybridization in specific cases, notably for `Rec`

Table 3. Results of pure news recommender systems in isolation and integrated with stochastic models across dimensions, part 2.

		Discovery (%)			*Time (Seconds)*		
NRS	**SM**	**AUS**	**Unexp**	**PCL**	**AccT**	**RecT**	**UpdT**
Pop		0.001±0.00	0.00±0.00	**100.0±0.00**	**20e-7±0.00**	87e-5±57e-6	**87e-3±61e-4**
Pop	CatF	▲**0.228±0.75**	△**0.027±0.09**	**100.0±0.00**	▲70e-7±18e-7	▼**25e-5±63e-6**	△16e-2±46e-3
Pop	Indep	▲0.047±0.12	△0.012±0.04	**100.0±0.00**	▲82e-7±20e-7	▼51e-5±12e-5	△16e-2±50e-3
Pop	Mk3	▲0.132±0.41	△0.012±0.04	**100.0±0.00**	▲60e-7±0.00	▼27e-5±40e-6	▲18e-2±34e-4
Pop	Mk4	▲0.136±0.42	△0.012±0.04	**100.0±0.00**	▲60e-7±21e-8	▼27e-5±42e-6	▲10e-1±23e-3
Pop	NV	▲0.138±0.40	△0.011±0.03	**100.0±0.00**	▲50e-6±14e-6	▼48e-5±12e-5	△16e-2±49e-3
Pop	RAD	▲0.120±0.37	△0.012±0.04	**100.0±0.00**	▲24e-5±27e-5	▼86e-5±54e-5	△16e-2±38e-3
Pop	CVD	▲0.095±0.29	△0.012±0.04	**100.0±0.00**	▲43e-6±12e-6	▼49e-5±12e-5	▽15e-2±36e-3
Rec		**100.0±0.00**	**98.082±1.99**	**100.0±0.00**	**50e-7±0.00**	**90e-7±0.00**	**81e-7±24e-7**
Rec	CatF	▼98.539±3.64	▼96.735±3.79	**100.0±0.00**	▲11e-6±73e-8	▲10e-5±84e-7	▲10e-5±35e-6
Rec	Indep	▼99.334±0.44	▼97.667±1.82	**100.0±0.00**	▲13e-6±84e-8	▲12e-5±65e-7	▲45e-5±14e-5
Rec	Mk3	▼98.971±1.48	▼97.197±2.14	**100.0±0.00**	▲75e-7±51e-8	▲79e-6±38e-7	▲98e-3±24e-4
Rec	Mk4	▼98.825±1.47	▼97.069±2.14	**100.0±0.00**	▲76e-7±49e-8	▲83e-6±50e-7	▲93e-2±32e-3
Rec	NV	▼99.036±1.56	▼97.220±2.21	**100.0±0.00**	▲65e-6±48e-7	▲17e-5±10e-6	▲85e-4±23e-4
Rec	RAD	▼99.030±1.49	▼97.256±2.15	**100.0±0.00**	▲36e-5±49e-5	▲76e-5±94e-5	▲87e-4±23e-4
Rec	CVD	▼99.055±1.36	▼97.329±2.06	**100.0±0.00**	▲58e-6±35e-7	▲15e-5±88e-7	▲79e-4±21e-4
Coo		20.624±16.47	**16.155±14.01**	**98.184±2.12**	**30e-7±21e-8**	**23e-5±95e-6**	**22e-7±85e-8**
Coo	CatF	▲**21.309±14.24**	▽14.604±11.22	▼95.571±4.18	▲80e-7±19e-7	▼40e-5±18e-5	▲85e-7±26e-7
Coo	Indep	▽20.224±15.62	▽14.690±13.36	▼96.231±3.59	▲85e-7±22e-7	▽45e-5±20e-5	▲27e-5±93e-6
Coo	Mk3	▲21.092±15.23	▽15.360±12.64	▼96.770±3.21	▲62e-7±43e-8	▲27e-5±80e-6	▲94e-3±21e-4
Coo	Mk4	▲21.194±15.27	▽15.418±12.67	▼96.747±3.23	▲65e-7±51e-8	▲27e-5±81e-6	▲90e-2±27e-3
Coo	NV	▲21.110±15.40	▽15.272±12.75	▼96.828±3.16	▲50e-6±13e-6	△46e-5±20e-5	▲55e-4±16e-4
Coo	RAD	△20.892±15.16	▽15.243±12.63	▼96.866±3.15	▲23e-5±27e-5	▲83e-5±45e-5	▲57e-4±17e-4
Coo	CVD	▽20.610±14.90	▽15.111±12.48	▼96.681±3.28	▲43e-6±12e-6	▽45e-5±20e-5	▲54e-4±17e-4
ICF		31.430±17.89	18.576±10.75	**88.988±9.03**	**45e-7±11e-7**	**30e-6±42e-7**	97.93±9.59
ICF	CatF	▽31.055±17.31	▽17.895±10.88	**88.988±9.03**	▲48e-7±80e-8	▲94e-6±63e-7	△90.22±15.32
ICF	Indep	▲**35.676±17.60**	▲**22.667±10.04**	▼88.567±9.11	▲53e-7±48e-8	▲17e-5±14e-6	▽**89.39±15.02**
ICF	Mk3	▲32.700±17.70	▲19.445±10.42	▼88.928±9.04	▲71e-7±81e-8	▲15e-5±12e-6	▽96.59±8.43
ICF	Mk4	▲32.744±17.71	▲19.455±10.42	▼88.908±9.04	▲73e-7±83e-8	▲15e-5±94e-7	▽97.16±8.67
ICF	NV	▲33.216±17.72	▲19.832±10.29	▼88.874±9.05	▲31e-6±20e-7	▲16e-5±99e-7	▽89.74±15.69
ICF	RAD	▲32.773±17.65	▲19.520±10.35	▼88.945±9.04	▲17e-5±24e-5	▲44e-5±46e-5	▽89.46±15.19
ICF	CVD	▲33.089±17.55	▲20.039±10.38	▼88.928±9.04	▲27e-6±13e-7	▲15e-5±95e-7	△90.07±15.15

and ICF. Regarding ILD, contextualization produces larger improvements than pure hybridization in systems such as Coo and ICF. In ELD and Pers, contextualization consistently outperforms hybridization across all NRS. For EPD, each NRS exhibits at least one configuration in which contextualization yields greater gains than hybridization. Finally, for the discovery metrics AUS and Unexp, this superiority is observed in Rec, Coo, and ICF, though not in Pop. Overall, the results indicate that stochastic models contextualization provides broader and more consistent improvements across diversity and personalization dimensions, while hybridization offers strong but more localized gains, especially in accuracy-related metrics.

When comparing hybrid systems with and without stochastic contextualization (Tables 4 and 5), we observe that combining both approaches yields greater gains, primarily in the diversity and personalization dimensions. In the diversity dimension, NRS+SM improves ILD in 2 of the 4 hybrid recommenders and ELD in all 4 cases. In the personalization, improvements are observed in 3/4 cases for Pers. In contrast, in the accuracy–coverage dimension, HRU shows no advantages, and the CC metric shows superiority in only 2 of the 4 hybrid recommenders. In the discovery dimension, there are no

Table 4. Results of hybrid news recommender systems in isolation and integrated with stochastic models across dimensions, part 1.

		Accuracy-Coverage (%)		*Diversity (%)*		*Novelty-Personalization (%)*	
NRS	**SM**	**HRU**	**CC**	**ILD**	**ELD**	**EPD**	**Pers**
Pop_H		**54.464**±**13.06**	0.004±95e-7	**85.593**±**8.24**	19.504±4.54	**21.543**±**5.57**	36.395±6.04
Pop_H	CatF	▼49.812±9.94	▲0.037±0.01	▼46.273±4.10	▲70.130±9.61	▼8.998±3.67	▲69.974±11.36
Pop_H	Indep	▼45.215±10.77	▲0.046±0.02	▽82.019±2.49	▲66.495±2.06	▽19.088±7.54	▲67.352±2.27
Pop_H	Mk3	▽53.401±9.56	▲0.056±0.02	▼71.995±5.02	▲72.073±5.38	▼10.708±4.40	▲70.825±5.44
Pop_H	Mk4	▽53.430±9.54	▲0.057±0.02	▼72.005±5.11	▲**72.362**±**5.53**	▼10.451±4.31	▲70.873±5.48
Pop_H	NV	▽52.956±9.60	▲0.055±0.02	▼73.504±4.77	▲71.600±5.07	▼10.411±4.40	▲71.067±4.70
Pop_H	RAD	▽52.806±9.67	▲**0.058**±**0.02**	▼73.533±4.48	▲71.992±4.50	▼11.116±4.82	▲**71.267**±**4.54**
Pop_H	CVD	▽51.989±9.91	▲**0.058**±**0.02**	▼75.537±4.02	▲71.706±3.42	▼12.675±5.03	▲71.221±4.26
Rec_H		**36.459**±**18.51**	0.003±0.00	**84.822**±**9.62**	13.019±6.69	**34.050**±**11.37**	32.324±10.38
Rec_H	CatF	▽32.090±14.31	▲0.029±0.01	▼52.531±7.78	▲65.841±13.06	▼18.147±12.48	▲68.705±10.46
Rec_H	Indep	▼28.432±14.82	▲0.034±0.03	▽82.684±4.74	▲63.717±5.55	▼24.587±11.58	▲64.999±4.05
Rec_H	Mk3	▽34.473±15.73	▲0.051±0.04	▼74.450±6.23	▲68.558±8.64	▼18.120±12.23	▲69.244±5.00
Rec_H	Mk4	▽34.523±15.73	▲**0.055**±**0.05**	▼74.441±6.31	▲**68.864**±**8.76**	▼18.007±12.26	▲69.317±5.02
Rec_H	NV	▽34.255±15.65	▲0.043±0.03	▼75.543±5.87	▲68.137±8.36	▼17.811±12.28	▲69.418±4.56
Rec_H	RAD	▽34.109±15.61	▲0.046±0.04	▼75.628±5.71	▲68.479±8.01	▼18.333±12.21	▲**69.545**±**4.46**
Rec_H	CVD	▼33.401±15.57	▲0.046±0.04	▼77.334±5.39	▲68.205±7.38	▼19.570±11.88	▲69.274±4.38
Coo_H		**64.987**±**6.07**	**6.910**±**2.16**	65.953±5.74	87.599±6.31	11.039±5.51	80.428±8.85
Coo_H	CatF	▼55.670±5.60	▼4.713±1.30	▼44.383±2.01	▲91.059±5.34	▼6.030±5.40	▲85.257±7.79
Coo_H	Indep	▼48.613±6.64	▼5.014±1.38	▲**82.006**±**1.65**	▲91.004±2.43	▲**18.693**±**8.27**	▲85.790±3.50
Coo_H	Mk3	▼57.517±5.30	▼5.648±1.61	▲71.298±3.67	▲92.287±3.51	▽10.001±5.76	▲86.876±4.97
Coo_H	Mk4	▼57.497±5.31	▼5.661±1.62	▲71.311±3.79	▲**92.344**±**3.50**	▽9.842±5.66	▲86.893±4.97
Coo_H	NV	▼56.903±5.36	▼5.681±1.65	▲72.990±3.52	▲92.151±3.40	▽9.352±5.03	▲87.106±4.56
Coo_H	RAD	▼56.959±5.34	▼5.662±1.61	▲72.867±3.16	▲92.330±3.30	▽10.091±5.39	▲**87.154**±**4.53**
Coo_H	CVD	▼56.474±5.43	▼5.433±1.47	▲74.906±2.81	▲92.282±3.04	△11.925±5.54	▲87.095±4.46
ICF_H		**52.296**±**6.63**	**3.441**±**0.51**	63.604±4.52	88.151±5.02	10.444±5.76	**84.714**±**7.56**
ICF_H	CatF	▼44.412±4.85	▼3.170±0.52	▼44.442±2.29	△**89.015**±**5.47**	▼5.464±4.51	▽84.181±7.79
ICF_H	Indep	▼34.130±6.07	▼2.341±0.47	▲**81.858**±**1.84**	▼81.675±2.07	▲**18.584**±**8.00**	▼78.293±3.04
ICF_H	Mk3	▼43.506±5.64	▼2.929±0.55	▲71.090±3.77	▽87.771±3.31	▽9.455±5.39	▼82.590±4.90
ICF_H	Mk4	▼43.524±5.65	▼2.914±0.55	▲71.091±3.89	▽87.900±3.33	▽9.268±5.30	▼82.622±4.91
ICF_H	NV	▼42.509±5.62	▼2.842±0.52	▲72.772±3.62	▽87.071±3.16	▽8.961±4.76	▼82.452±4.48
ICF_H	RAD	▼42.322±5.64	▼2.980±0.55	▲72.703±3.25	▽87.577±3.03	▽9.725±5.18	▽82.578±4.47
ICF_H	CVD	▼41.302±5.82	▼2.903±0.57	▲74.772±2.90	▽86.943±2.68	△11.660±5.45	▼82.235±4.41

gains in `AUS`, while there is an advantage in 2/4 cases in `Unexp`; moreover, contextualization does not reintroduce cold-start issues previously resolved by hybridization, maintaining $PCL = 1$. Finally, as in the non-hybrid scenario, SM add processing time, but the values remain very small and do not change the computational cost scale of the systems.

Comparing our five SM with the two adversarial baselines in the hybrid scenario reveals meaningful shifts across evaluation dimensions. In the accuracy–coverage dimension, our models outperform `CatF`, reaching the Pareto frontier in 3 out of 4 cases. In the diversity and novelty–personalization dimensions, they remain tied with `Ind`, achieving the same number of frontier positions while clearly surpassing `CatF`, which rarely reaches the frontier in these metrics. In the discovery dimension, performance becomes more balanced: our models achieve frontier positions in 2 out of 4 cases, compared to 3 out of 4 for `CatF`, indicating a slight advantage for adversarial in this specific dimension.

By incorporating user behavior modeling, stochastic contextualization enhances personalization and diversity across all base systems, regardless of their intrinsic characteristics. In simple, non-personalized recommenders, such as popularity- or recency-based systems, the stochastic models produce substantial gains in accuracy, coverage, di-

Table 5. Results of hybrid news recommender systems in isolation and integrated with stochastic models across dimensions, part 2.

		Discovery (%)			*Time (Seconds)*		
NRS	**SM**	**AUS**	**Unexp**	**PCL**	**AccT**	**RecT**	**UpdT**
Pop_H		**24.997±28.00**	**18.603±22.16**	**100.0±0.00**	**22e-7±39e-8**	25e-4±13e-5	**25e-2±15e-3**
Pop_H	CatF	▽22.634±21.00	▽17.958±16.79	**100.0±0.00**	▲63e-7±17e-7	▼**40e-5±12e-5**	▼31e-2±67e-3
Pop_H	Indep	▽22.434±24.26	▽18.047±20.48	**100.0±0.00**	▲73e-7±22e-7	▼92e-5±24e-5	▽33e-2±74e-3
Pop_H	Mk3	▽23.642±23.37	▽18.329±18.80	**100.0±0.00**	▲61e-7±29e-8	▼56e-5±11e-5	▲32e-2±12e-3
Pop_H	Mk4	▽23.728±23.29	▽18.383±18.73	**100.0±0.00**	▲61e-7±29e-8	▼56e-5±11e-5	▲1.13±15e-3
Pop_H	NV	▽23.813±23.53	▽18.040±18.72	**100.0±0.00**	▲44e-6±14e-6	▼81e-5±24e-5	▽33e-2±91e-3
Pop_H	RAD	▽23.502±23.43	▽18.030±18.77	**100.0±0.00**	▲22e-5±27e-5	▼12e-4±62e-5	▼31e-2±83e-3
Pop_H	CVD	▽23.137±23.41	▽18.177±19.00	**100.0±0.00**	▲37e-6±12e-6	▼84e-5±22e-5	▼31e-2±72e-3
Rec_H		18.165±29.70	14.682±27.57	**100.0±0.00**	**20e-7±0.00**	**51e-5±27e-6**	**52e-3±12e-3**
Rec_H	CatF	▲**28.448±27.97**	▲**25.015±26.84**	**100.0±0.00**	▲55e-7±19e-7	▲78e-5±29e-5	▲69e-2±32e-2
Rec_H	Indep	▲23.308±28.82	▲20.701±27.45	**100.0±0.00**	▲67e-7±23e-7	▲19e-4±77e-5	▲66e-2±26e-2
Rec_H	Mk3	▲25.842±28.67	▲22.412±27.29	**100.0±0.00**	▲62e-7±39e-8	▲11e-4±74e-6	▲58e-2±23e-3
Rec_H	Mk4	▲26.189±28.61	▲22.743±27.23	**100.0±0.00**	▲63e-7±48e-8	▲11e-4±78e-6	▲1.38±35e-3
Rec_H	NV	▲25.543±28.71	▲21.868±27.29	**100.0±0.00**	▲42e-6±16e-6	▲16e-4±64e-5	▲67e-2±25e-2
Rec_H	RAD	▲25.250±28.72	▲21.801±27.31	**100.0±0.00**	▲22e-5±27e-5	▲20e-4±83e-5	▲68e-2±26e-2
Rec_H	CVD	▲24.654±28.74	▲21.568±27.34	**100.0±0.00**	▲36e-6±14e-6	▲17e-4±68e-5	▲66e-2±24e-2
Coo_H		**17.228±16.53**	14.348±13.83	**100.0±0.00**	**51e-7±35e-8**	**39e-5±98e-6**	**22e-2±14e-3**
Coo_H	CatF	▽17.097±14.09	△**14.556±11.94**	**100.0±0.00**	▲10e-6±24e-7	▲79e-5±34e-5	▽38e-2±85e-3
Coo_H	Indep	▽15.216±16.04	▽13.060±14.02	**100.0±0.00**	▲11e-6±26e-7	▲10e-4±38e-5	△39e-2±84e-3
Coo_H	Mk3	▽16.773±15.36	▽13.994±12.98	**100.0±0.00**	▲80e-7±21e-8	▲59e-5±17e-5	▲31e-2±12e-3
Coo_H	Mk4	▽16.921±15.35	▽14.132±12.99	**100.0±0.00**	▲80e-7±21e-8	▲60e-5±18e-5	▲1.13±18e-3
Coo_H	NV	▽16.761±15.49	▽13.775±12.98	**100.0±0.00**	▲50e-6±14e-6	▲94e-5±38e-5	▽37e-2±78e-3
Coo_H	RAD	▽16.537±15.27	▽13.718±12.89	**100.0±0.00**	▲23e-5±27e-5	▲13e-4±73e-5	▽37e-2±77e-3
Coo_H	CVD	▽16.020±15.15	▽13.518±12.89	**100.0±0.00**	▲45e-6±13e-6	▲94e-5±37e-5	▽36e-2±59e-3
ICF_H		**23.618±17.65**	**20.671±14.87**	**100.0±0.00**	**41e-7±35e-8**	**34e-5±27e-5**	84.91±16.92
ICF_H	CatF	▽22.977±16.95	▽19.911±14.67	**100.0±0.00**	▲65e-7±74e-8	▲11e-4±28e-5	△81.65±16.62
ICF_H	Indep	▽22.153±20.51	▽19.236±18.01	**100.0±0.00**	▲72e-7±91e-8	▲28e-4±19e-5	△82.05±16.48
ICF_H	Mk3	▽22.999±19.21	▽19.444±16.38	**100.0±0.00**	▲13e-6±42e-7	▲23e-4±30e-5	△85.53±16.51
ICF_H	Mk4	▽23.193±19.14	▽19.617±16.31	**100.0±0.00**	▲12e-6±39e-7	▲23e-4±31e-5	▲86.23±16.59
ICF_H	NV	▽23.244±19.50	▽19.240±16.37	**100.0±0.00**	▲33e-6±25e-7	▲23e-4±19e-5	▽**81.35±16.24**
ICF_H	RAD	▽22.850±19.33	▽19.181±16.41	**100.0±0.00**	▲18e-5±26e-5	▲25e-4±61e-5	▽81.48±16.22
ICF_H	CVD	▽22.472±19.40	▽19.248±16.68	**100.0±0.00**	▲28e-6±18e-7	▲23e-4±18e-5	▽81.40±16.06

versity, and personalization by contextualization of the recommendations, although novelty tends to decrease as recommendations become more aligned with user behavior. In stronger personalized recommenders, such as co-occurrence and item-based collaborative filtering, the stochastic models reinforce user-level differentiation and inter-user diversity while introducing only minor reductions in accuracy. These improvements arise from stochastic models' ability to capture temporal structure, contextual dependencies, and behavioral regularities in users' reading trajectories, dimensions that traditional NRS typically underexplores. By encoding sequential patterns observed in real consumption flows, stochastic contextualization steers recommendations toward content that is simultaneously coherent with short-term intent and more diverse within and across users. This mechanism ultimately yields recommendations that are more personalized, less redundant, and more behavior-aware.

The incorporation of SM slightly increases computational complexity but preserves the high efficiency of all systems. Response times (`AccT` + `RecT`) remain well below one second, confirming suitability for real-time deployment. Although `ICF` originally exhibits the highest update cost, stochastic contextualization does not amplify this behavior, and because the system performs updates asynchronously, online responsiveness remains unaffected. Overall, the inclusion of SM maintains the required computational performance while delivering measurable gains across multiple dimensions.

Our work has some limitations: first, we did not aim to fine-tune the baselines with optimal parameters or to propose new recommendation algorithms. Similarly, we did not attempt to generate more accurate recommendations. Due to the lack of compatible framework implementations, we were unable to test incorporating stochastic models into pre-existing systems and instead implemented the systems ourselves, following the literature. We have not evaluated the integration of stochastic models alongside other approaches, such as content-based recommender systems. We also limited our experiments to match the configuration of the real news site from which we collected the data. Our goal was not to find the optimal list size, but to simulate a realistic recommendation scenario under the same constraints. Finally, while the dataset underwent basic cleaning and preparation (removing duplicates and incomplete data) to facilitate sequential access analysis, we did not filter user or item accesses to improve recommendation results artificially. Accordingly, our evaluation reflects a realistic deployment scenario, in which recommendations would be generated regardless of access type, without additional filters or constraints.

6. Conclusion

This paper explored integrating stochastic models into news recommender systems to enhance effectiveness beyond traditional approaches. Stochastic models amplify the strengths and reveal the weaknesses of the base recommender. When the data are generic and lack personalization, stochastic modeling acts as a contextual filter, aligning recommendations with user profiles and improving diversity, and personalization. Conversely, in systems well-tuned to data but constrained by structural issues such as cold-start, stochastic models may exacerbate these problems by further restricting the recommendation space. Ultimately, the effectiveness of stochastic modeling depends on the quality and nature of the underlying system.

As future work, we propose exploring the impact of varying parameters, such as the number of categories, category granularity, the order of Markov models, and the length of user history, on the performance of stochastic contextualization. Additional directions include developing personalized stochastic models for recurring users, investigating alternative model classes and integration strategies, and extending the evaluation to other news recommender systems. Future studies should also consider systematic hyperparameter optimization of all compared recommenders to strengthen the fairness of comparisons, as well as experiments on multiple news datasets and platforms to further assess the generalizability of the observed effects.

Acknowledgment. The authors express their gratitude to the Federal University of Ouro Preto for its support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, CNPq (306572/2019-2), and INCT-TILDIAR (grant # 408490/2024-1).

In preparing this paper, we used Grammarly and ChatGPT to identify and correct grammatical errors, typos, and other writing mistakes.

References

Bai, X., Cambazoglu, Gullo, F., Mantrach, A., and Silvestri, F. (2017). Exploiting search history of users for news personalization. *Information Sciences*, 385-386:125–137.

Campos, P. G., Díez, F., and Cantador, I. (2014). Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1-2):67–119.

Çano, E. and Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21(6):1487–1524.

Cavalcante, P. S. and Pinheiro, W. A. (2013). Mecanismo de encadeamento de notícias por reconhecimento de implicação textual. In *SBBD (Short Papers)*, pages 27–1.

Champiri, Z. D., Shahamiri, and Salim (2015). A systematic review of scholar context-aware recommender systems. *Expert Systems with Applications*, 42(3):1743–1758.

Constantinides, M. and Dowell, J. (2018). A framework for interaction-driven user modeling of mobile news reading behaviour. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, page 33–41. ACM.

Das, A. S., Datar, M., Garg, A., and Rajaram, S. (2007). Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web*, page 271–280. ACM.

Epure, E. V., Kille, B., Ingvaldsen, J. E., Deneckere, R., Salinesi, C., and Albayrak, S. (2017). Recommending personalized news in short user sessions. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, page 121–129. ACM.

Garcin, F., Dimitrakakis, C., and Faltings, B. (2013). Personalized news recommendation with context trees. In *RecSys 2013 - Proceedings of the 7th ACM Conference on Recommender Systems*, pages 105–112. ACM.

Ji, Z., Wu, M., Yang, H., and Íñigo, J. E. A. (2021). Temporal sensitive heterogeneous graph neural network for news recommendation. *Future Generation Computer Systems*, 125:324–333.

Karimi, M., Jannach, D., and Jugovac, M. (2018). News recommender systems – survey and roads ahead. *Information Processing & Management*, 54(6):1203–1227.

Koo, B., Jeon, H., and Kang, U. (2021). Pgt: news recommendation coalescing personal and global temporal preferences. *Knowledge and Information Systems*, 63:3139–3158.

Li, J., Zhu, J., Bi, Q., Cai, G., Shang, L., Dong, Z., Jiang, X., and Liu, Q. (2022). MINER: Multi-interest matching network for news recommendation. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 343–352. ACL.

Li, L., Zheng, L., Yang, F., and Li, T. (2014). Modeling and broadening temporal user interest in personalized news recommendation. *Expert Systems with Applications*, 41:3168–3177.

Liu, J., Dolan, P., and Pedersen, E. R. (2010). Personalized news recommendation based on click behavior. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, pages 31–40. ACM.

Lv, P., Zhang, Q., Shi, L., Guan, Z., Fan, Y., Li, J., Zhong, K., and Deveci, M. (2024). Exploring on role of location in intelligent news recommendation from data analysis perspective. *Information Sciences*, 662.

Maksai, A., Garcin, F., and Faltings, B. (2015). Predicting online performance of news recommender systems through richer evaluation metrics. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 179–186. ACM.

Meng, Q., Yan, H., Liu, B., Sun, X., Hu, M., and Cao, J. (2023). Recognize news transition from collective behavior for news recommendation. *ACM Transactions on Information Systems*, 41.

Moreira, G. D. S. P., Ferreira, F., and Cunha, A. M. D. (2018). News session-based recommendations using deep neural networks. In *ACM International Conference Proceeding Series*, pages 15–23. ACM.

Park, C. (2025). Llm as user simulator: Towards training news recommender without real user interactions. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 3080–3084. ACM.

Pu, Z. and Beam, M. A. (2024). The impacts of relevance of recommendations and goal commitment on user experience in news recommender design. *User Modeling and User-Adapted Interaction*, 34(4):925–953.

Raza, S. and Ding, C. (2022). News recommender system: a review of recent progress, challenges, and opportunities. *Artif. Intell. Rev.*, 55(1):749–800.

Ren, S. and Shi, C. (2022). A news recommendation model based on time awareness and news relevance. In *23rd International Conference on Information Reuse and Integration for Data Science (IRI)*, page 35–40, USA. IEEE Press.

Sheu, H. S., Chu, Z., Qi, D., and Li, S. (2022). Knowledge-guided article embedding refinement for session-based news recommendation. *IEEE Transactions on Neural Networks and Learning Systems*, 33:7921–7927.

Symeonidis, P., Chaltsev, D., Berbague, C., and Zanker, M. (2022). Sequence-aware news recommendations by combining intra- with inter-session user information. *Information Retrieval Journal*, 25:461–480.

Veloso, B. M., Assunção, R. M., Ferreira, A. A., and Ziviani, N. (2019). In search of a stochastic model for the e-news reader. *ACM Transactions on Knowledge Discovery from Data*, 13(6):1–27.

Wang, Y., Sang, L., Zhang, Y., and Zhang, Y. (2025). Intent representation learning with large language model for recommendation. In *Proceedings of the 48th International ACM SIGIR*, page 1870–1879. ACM.

Wu, C., Wu, F., Huang, Y., and Xie, X. (2023). Personalized news recommendation: Methods and challenges. *ACM Trans. Inf. Syst.*, 41(1).

Xin, R., Chen, X., Jiang, D., He, Y., Ou, Z., Liu, P., Han, Z., and Song, M. (2022). El-rec: Enhanced user and news interaction for news recommendation. In *8th International Conference on Cloud Computing and Intelligent Systems*, pages 355–359. IEEE.