Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 3, 2026, 311-327.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

DOTA-Draft: A Dataset for In-Game Recommendation in Multiplayer Online Battle Arenas

Mohammadreza Mohammadnejad, Morteza Dorrigiv* and Farzin Yaghmaee

Electrical and Computer Engineering Department, Semnan University, Semnan, Iran.

Article Info

Article History:

Received 04 October 2025
Revised 19 December 2025
Accepted 20 May 2026

*DOI:*10.22044/jadm.2026.16888.2819

Keywords:

In-game Recommendation Systems, Multiplayer Online Battle Arenas (MOBA), Dota 2 Interaction Dataset, Session-based Recommendation Models, Benchmarking with RecBole Framework.

**Corresponding author: dorrigiv@semnan.ac.ir (M. Dorrigiv).*

Abstract

Research in recommender systems has largely relied on standardized datasets such as MovieLens, Amazon Reviews, and Last.fm. However, these datasets are unsuitable for in-game recommendations, particularly in Multiplayer Online Battle Arenas (MOBAs), due to the sequential, team-based, and adversarial nature of gameplay. To identify essential characteristics for in-game recommendation datasets, we perform a cross-domain analysis of widely used recommendation datasets, evaluating their structural and distributional properties, including interaction space, matrix shape, sparsity, and Gini-based feature–shape diversity. Building on these insights, we curate DOTA-Draft, a research-ready dataset from raw professional Dota 2 matches, encoding sequential pick/ban states, patch versions, and match outcomes. Using this dataset, we conduct top-k drafting recommendation tasks and provide baseline results with Bayesian Personalized Ranking (BPR) and GRU4Rec. To facilitate adoption, DOTA-Draft is packaged in a RecBole-compatible format. This work establishes principled benchmarks for in-game recommendation, demonstrates the inadequacy of traditional user–item paradigms in dynamic, adversarial environments, and provides a foundation for developing models that account for sequential, multi-agent decision-making.

1. Introduction

Recommender systems have become an essential component of modern digital platforms, shaping personalized experiences in e-commerce, entertainment, and social media. Their success is largely attributed to the availability of high-quality, standardized datasets, such as MovieLens for movies [1] and Amazon Reviews for e-commerce [2,3,4]. These resources have been instrumental in benchmarking, ensuring reproducibility, and accelerating advancements in the field.

Over the past decade, the digital entertainment industry—particularly video games—has grown into a multi-billion-dollar market [5]. This growth has drawn significant academic attention across diverse areas, including behavior-analysis [6,7,8] and gender bias detection [9,10]. Within the software engineering (SE) community, games have also been recognized as valuable software artifacts [11], opening new research avenues such as code smell detection [12], sentiment analysis [13,14], and serious games [15].

However, a persistent challenge in this domain lies in the difficulty of data collection. Unlike other domains with well-established datasets, many digital games lack proper APIs or provide unstructured data, requiring substantial development effort to collect and prepare information for research [16]. This limits reproducibility and constrains the development of robust in-game recommendation models.

In this context, in-game recommender systems—especially in complex, multi-agent environments

like MOBAs—remain relatively underexplored. Despite the growing significance of games as one of the largest sectors in digital entertainment, the absence of public, standardized datasets has constrained research progress. In-game recommendation tasks, such as drafting heroes, recommending in-game items, or suggesting strategies, present unique challenges that differ fundamentally from those in other domains.

Decision-making in MOBA environments is characterized by strong sequential and temporal dynamics, where choices are made in real time. These decisions evolve across game patches, balance updates, and shifting player strategies [17,18]. In contrast to individual preference domains such as movie recommendation, MOBA drafting and gameplay require modeling complex team-based interactions, including synergy and counter-synergy among multiple agents [18]. Moreover, these environments exhibit high feature diversity, as contextual factors span heroes, roles, items, player skill levels, and patch versions, resulting in a large, heterogeneous, and continually evolving feature space [18,19]. Existing datasets are generally designed for game recommendation to users (e.g., which game a player might enjoy) [20,21,22] rather than in-game recommendation tasks. Even when raw data is available—for example, logs of professional Dota 2 matches on Kaggle [23]—it is often not research-ready: typically stored in flat CSV files, lacking standardized schemas, and unsuitable for direct use in recommender frameworks. For studies that do focus on in-game recommendations [24,25,26,27,28,29,30,31], most rely on APIs to collect raw data. However, this data usually requires extensive preprocessing and normalization, making reproducibility difficult and limiting the ability to build upon prior work. This reveals a critical gap for the recommender systems community.

To address this gap, we make the following contributions:

1. Cross-Domain Dataset Analysis: We analyze research-ready datasets from multiple domains (e.g., movies, e-commerce, music, gaming) and examine their features (e.g., shape, space, diversity, Gini for items, Gini for users), demonstrating the absence of suitable benchmarks for in-game recommendation research.
2. Dataset Construction: We preprocess and structure raw professional Dota 2 match data from Kaggle [23] into research-ready datasets (i.e., DOTA-Draft), designed for drafting recommendations by encoding sequential pick/ban states along with patch and outcome labels.
3. Benchmark Packaging: We package DOTA-Draft for the RecBole [32,33] recommendation framework and provide baseline evaluations using standard Bayesian Personalized Ranking (BPR) [34] and GRU4Rec.
4. Open Challenges: We highlight the unique challenges of in-game recommendation—such as patch-driven dynamics, multi-agent interactions, and high-dimensional context, which demand new algorithms beyond those used in traditional recommendation tasks.

By releasing these datasets and benchmarks, we aim to speed up research in in-game recommendation, bridging the gap between the recommender systems community and the fast-growing domain of online games.

To provide a comprehensive understanding of dataset creation for in-game recommendation systems, the remainder of this paper is organized as follows. Section 2 reviews related work in the fields of recommender systems and gaming datasets, highlighting existing approaches and their limitations. Section 3 presents a survey of current datasets and conducts a gap analysis to identify unmet needs in the domain. Section 4 introduces our proposed benchmark dataset, detailing its design principles, features, and intended use cases. Section 5 describes the benchmarking and evaluation process, including experimental setup and performance metrics. Section 6 discusses the implications of our findings, potential applications, and limitations. Finally, Section 7 concludes the paper and outlines directions for future research.

2. Related Works

Recommendation research in MOBAs has primarily focused on Dota 2, with studies addressing hero drafting, match prediction, and item usage. However, unlike mature recommendation domains, this line of work lacks standardized, research-ready datasets, forcing reliance on ad hoc data collection or heavily preprocessed game logs.

Early work [18] explored hero drafting by leveraging the Steam API and applying association rules combined with neural networks. While effective in simulating drafting phases, the study relied on custom API crawls and did not release a benchmark-ready dataset, limiting reproducibility and reuse.

Similarly, prior work [25] evaluated classical machine learning approaches for hero

recommendation and outcome prediction using data collected via the Steam Web API. However, the resulting dataset was not released in a standardized or reusable form, limiting reproducibility and comparative evaluation.

Beyond individual hero selection, research has also addressed structural fairness in matchmaking. For instance, the CUPID framework [35] introduces a re-matchmaking system that optimizes team and position assignments to improve fairness and player satisfaction. While it leverages large-scale, real-world MOBA data to achieve high win-prediction accuracy, the dataset is proprietary, tied to a specific collection window, and specialized for matchmaking rather than hero-drafting recommendation benchmarks.

Replay files have also been explored as a richer data source. For example, the study "Creating a Hero Recommender System for Newer Players in Dota 2" [28] processed 3,483 high-skill replays into CSV files, capturing detailed hero and item events. While this enabled fine-grained analysis, the resulting dataset was tailored to a single thesis project and not standardized for broader recommender system research.

More recent efforts have focused on real-time decision support. Blaszczyk and Szajerman [36] evaluated ensemble models for League of Legends during the live pick-and-ban phase, showing that player-specific statistics dominate outcome prediction. However, their datasets—derived from Riot API crawls and Kaggle archives—remain task-specific and fragmented, reinforcing the need for unified benchmarking frameworks. The complexity of hero relationships has led to graph-based architectures. Chen [37] proposed MOBARec-GCNFP, a lightweight GCN for counter-pick recommendation, addressing cold-start issues via embedding initialization. While achieving strong Recall and NDCG performance, such models depend on specialized graph construction and feature engineering, underscoring the need for standardized datasets that support graph-based and sequential evaluation. DraftRec [29] introduced a personalized drafting recommendation model evaluated on large-scale League of Legends and Dota 2 match archives. Although this work advanced algorithmic modeling, it again relied on third-party match archives not designed for recommender tasks, requiring significant preprocessing. The datasets were useful for experiments but not packaged for general reuse within frameworks like RecBole [32]. Across prior studies, a recurring issue is evident: datasets are often collected in fragmented ways, lack standardized schemas, and remain absent from established recommender benchmarks. This fragmentation forces researchers to duplicate data-cleaning efforts, hinders reproducibility, and complicates fair model comparison across studies. To address these challenges, we introduce DOTA-Draft, a research-ready Dota 2 dataset for drafting recommendation, explicitly packaged for reproducibility and seamless benchmarking within RecBole [32,33].

3. Dataset Survey and Gap Analysis

MOBAs such as Dota 2 and League of Legends involve strategic, sequential, and highly interactive decision-making, setting them apart from traditional consumption domains. Unlike recommending the next movie or product, in-game recommendations—such as suggesting the next hero pick or item—must account for evolving game states, player strategies, and competitive balance.

To identify gaps, we conducted a cross-domain survey of widely research-ready datasets. Our survey reveals a critical gap: although large-scale game data exist in raw forms (e.g., match logs, Kaggle dumps), there is no standardized, research-ready dataset for in-game recommendation. Data are often noisy, inconsistently formatted, and lack preprocessing pipelines or metadata required for reproducibility. Moreover, existing resources do not integrate with modern frameworks such as RecBole, limiting their usability for systematic benchmarking.

To ground our work in established practices, we first conducted a comparative study of widely used datasets from multiple domains. Figure 1 presents a treemap visualization of dataset distribution across categories, offering an overview of how datasets are spread over different application domains.

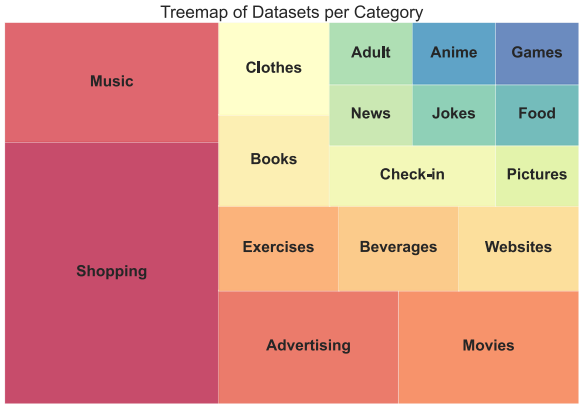

Figure 1. Treemap visualization of the dataset distribution across categories.

Complementing this, Figure 2 provides a bar chart summarizing the same distribution, highlighting both dominant categories and under-represented ones. Beyond raw counts, Table 1 reports detailed dataset statistics, including interaction counts, space–shape density, and measures of inequality such as user and item Gini coefficient. These visualizations and statistics jointly illustrate the breadth and diversity of existing benchmarks while exposing gaps relevant to in-game recommendation research.

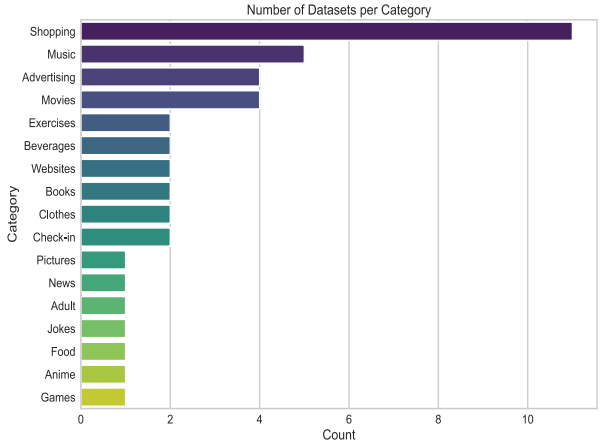

Figure 2. Bar chart of dataset distribution by category.

This gap motivates our work: the development of DOTA-Draft, a specialized Dota 2 benchmark dataset for drafting recommendation, explicitly designed to capture the sequential, contextual, and strategic nature of in-game decision-making.

Table 1 presents a collection of benchmark datasets commonly used in recommender system research, organized by application domain. Each row corresponds to a dataset, showing the number of users, items, and recorded interactions. Grouping by category highlights the diversity of domains from adult content and music to books and games. Notably, the datasets in the games category focus on recommending games to users rather than providing in-game recommendations, distinguishing them from other domains where recommendations occur within the application context.

This categorization enables systematic comparison between datasets, while the reported statistics provide insight into their relative scale and complexity.

3.1. Dataset Characteristics

Following prior work [38], recommendation datasets can be characterized along two dimensions: structural and distributional properties. These characteristics are derived directly from the user–item interaction matrix and provide insights into the scale, balance, and skew of a dataset.

- U denotes the set of users, with cardinality $|U|$
- I denotes the set of items, with cardinality $|I|$
- $K \subseteq U \times I$ denotes the set of observed user–item interactions, with cardinality $|K|$
- K_u denotes the set of interactions associated with user u
- K_i denotes the set of interactions associated with item i.

Structural characteristics — namely space [Eq. (1)], shape [Eq. (2)], and density [Eq. (3)] — are defined as follows:

- Space:

$$Space_{log} = \log_{10}\left(\frac{|U| \times |I|}{sc}\right) \quad (1)$$

- Shape:

$$Shape_{log} = log_{10}\left(\frac{|U|}{|I|}\right) \quad (2)$$

- Density:

$$\text{Density}_{log} = log_{10}\left(\frac{|K|}{|U| \times |I|}\right) \quad (3)$$

Distributional characteristics—namely the Gini-user [Eq. (4)] and the Gini-item [Eq. (5)]—quantify how interactions are distributed across users and items:

- Gini-user:

$$\text{Gini}_{\text{user}} = 1 - 2\sum_{u=1}^{|U|}\left(\frac{|U| + 1 - u}{|U| + 1}\right) \times \left(\frac{|K_u|}{|K|}\right) \quad (4)$$

- Gini-item:

$$\text{Gini}_{\text{item}} = 1 - 2\sum_{i=1}^{|I|}\left(\frac{|I| + 1 - i}{|I| + 1}\right) \times \left(\frac{|K_i|}{|K|}\right) \quad (5)$$

Table 1. Key Dataset Metrics and Categories.

Dataset	Users	Items	Interactions	$Space_{log}$	$Shape_{log}$	$Density_{log}$	$Gini_{User}$	$Gini_{Item}$	Category
Adult [39]	Not Available	Not Available	32561						Adult
AliEC [40]	491647	240130	1366056	11.072100	0.311207	-4.936631	4.14E-12	3.48E-12	Advertising
Avazu [41]	Not Available	Not Available	40428967						
iPinYou [42]	19731660	163	24637657	9.507351	5.082976	-2.115752	-2.82E-10	1.39E-15	
Criteo [40]	Not Available	Not Available	45850617						
Anime [32]	73515	11200	7813737	8.915594	0.817158	-2.022735	7.53E-13	1.65E-13	Anime
RateBeer [43]	29265	110369	2924163	9.509196	-0.576499	-3.043194	2.06E-13	9.22E-13	Beverages
BeerAdvocate [43]	33388	66055	1586614	9.343496	-0.296315	-3.143025	8.00E-14	5.86E-13	
GoodReads [21]	876145	2360650	228648342	12.315608	-0.430456	-3.956440	-3.60E-13	1.46E-11	Books
Book-Crossing [32]	105284	340557	1149780	10.554552	-0.509827	-4.493937	9.80E-13	2.87E-12	
Gowalla [44]	107092	1280969	6442892	11.137296	-1.077782	-4.328215	-7.09E-14	1.79E-11	Check-in
Foursquare [32]	Not Available	Not Available	227428						
RentTheRunway [45]	105571	5850	192544	8.790701	1.256389	-3.506171	8.34E-13	5.88E-14	Clothes
ModCloth [45]	47958	1378	82790	7.820110	1.541612	-2.902132	2.36E-13	1.30E-14	
EndoMondo [46]	1104	253020	253020	8.446124	-2.360186	-3.042969	9.89E-15	0	Exercises
KDD2010 [47]	Not Available	Not Available	20768884						
Food [48]	226570	231637	1132367	10.720010	-0.009606	-4.666023	-2.83E-12	-8.60E-13	Food
Steam [20]	2567538	32135	7793069	10.916495	1.902539	-4.024787	3.72E-11	1.42E-14	Games
Jester [32]	73421	101	4136360	6.870142	2.861499	-0.253523	5.95E-15	-1.41E-16	Jokes
Douban [49]	738701	28	2125056	7.315627	4.421311	-0.988256	5.32E-12	-3.81E-16	Movies
Twitch [50]	15524309	6161666	474676929	13.980710	0.401314	-5.304312	4.81E-13	6.67E-11	
MovieLens-20m [1]	138493	27278	20000263	9.577240	0.705615	-2.276205	-1.18E-12	2.40E-13	
Netflix [32]	480189	17770	100480507	9.931100	1.431725	-1.929018	-3.69E-12	-3.58E-14	
KGRec [51]	21552	20000	2117698	8.634518	-0.032458	-2.308654	-1.00E-13	-2.42E-13	Music
Last.FM [52]	1892	17632	92834	7.523223	-0.969380	-2.555516	1.17E-14	1.09E-13	
LFM-1b [53]	120322	3123496	1088161692	11.574986	-1.414296	-2.538293	1.44E-12	-3.47E-11	
Yahoo Music [32]	1948882	98211	111557943	11.281946	1.297625	-3.234445	-1.06E-11	-6.76E-13	
Music4All-Onion [54]	119140	109269	252984396	10.114555	0.037561	-1.711461	-1.07E-12	-1.75E-13	
MIND [55]	1000000	161013	24155470	11.206861	0.793139	-3.823845	-1.02E-11	2.31E-12	News
Pinterest [56]	55187	9911	1445622	8.737954	0.745719	-2.577900	-4.57E-14	-8.99E-14	Pictures
Epinions [57]	116260	41269	188478	9.681054	0.449806	-4.405794	9.55E-13	6.70E-15	Shopping
DianPing [58]	542706	243247	4422473	11.120612	0.348517	-4.474947	-2.02E-12	1.62E-12	
Amazon_M2 [32]	3606349	1410675	15306183	12.706495	0.407641	-5.521628	5.19E-12	1.89E-11	
Amazon [32]	54510000	48190000	571540000	15.419433	0.053519	-6.662386	-3.11E-10	6.13E-10	
Retailrocket [59]	1407580	247085	2756101	11.541319	0.755627	-5.101024	1.40E-11	-1.88E-12	
Alibaba-iFashion [32]	3569112	4463302	191394393	13.202216	-0.097096	-4.920287	2.42E-11	-4.23E-11	
Ta Feng [32]	32266	23812	817741	8.885541	0.131949	-2.972925	3.39E-13	-2.50E-13	
Tmall [32]	963923	2353207	44528127	12.355702	-0.387618	-4.707068	7.78E-12	-2.95E-11	
DIGINETICA [32]	204789	184047	993483	10.576235	0.046378	-4.579075	-7.33E-14	8.20E-13	
Yelp-full [32]	5556436	539254	28908420	12.476590	1.013003	-5.015565	2.86E-11	8.45E-13	
YOOCHOOSE [32]	9249729	52739	34154697	11.688261	2.243997	-4.154810	-3.41E-11	5.82E-14	
Phishing websites [60]	NAN	NAN	11055						Websites
Behance/Vista dataset [61]	63497	178788	1000000	10.055092	-0.449585	-4.055092	-1.15E-16	-2.30E-12	

$Gini_{user}$ and $Gini_{item}$ utilize the Gini coefficient to measure the distribution of interactions over the set of users U and the set of items I, respectively. Since these two characteristics are similar in nature, we will describe them by using $Gini_{item}$ as an example. $Gini_{item}$ takes values between 0 to 1 (inclusive), and a value of 0 implies total equality (i.e. all items are equally popular and have the same number of observed interactions), while a value of 1 indicates maximal inequality whereby all observed interactions belong to a single item. In reality, skewed frequency distributions are rather common across many application scenarios [38]. For example, a blockbuster movie can attract much more viewers as compared to another movie with a relatively smaller target audience.

4. Proposed Benchmark Dataset

In this section, we present research-ready datasets specifically designed for in-game recommendation in MOBA environments, with a focus on Dota 2. The dataset was constructed from raw professional-match data available on Kaggle [23], which, while rich in content, is stored in flat CSV files that are not directly usable for recommendation research. Our preprocessing pipeline transforms this raw data into structured benchmarks with well-defined schemas, tailored tasks, and compatibility with existing recommendation frameworks.

4.1. Data Source

The datasets are derived from publicly available logs of professional Dota 2 matches (in 2024) on Kaggle [23]. Each match contains information on participating teams, heroes, item purchases, bans, outcomes, and timestamps. Although extensive, the original CSV structure lacks the consistency, normalization, and formatting required for recommender systems research.

4.2. DOTA-Draft Dataset

The drafting stage of a MOBA match can be formulated as a sequential recommendation problem. In our dataset construction, we follow the real-world structure of MOBA drafts, which consist of 24 alternating ban and pick turns rather than the simplified ten-pick setting used in prior work [29].

Each match is represented as an ordered sequence of draft actions, where each row encodes the acting side, the action type (i.e., ban or pick), the selected hero, and contextual information such as patch version and match outcome. This formulation allows bans to be modeled as constraints that dynamically reduce the candidate hero pool and ensures that the asymmetric pick order and side outcomes are preserved. By aligning the dataset with the 24-step drafting phase used in esports tournaments, we provide a foundation for studying draft prediction and recommendation under realistic competitive conditions.

the dataset D is defined as in Eq. (6)

$$D = \left\{ \left(s_i^t, a^t, c_i^t, y_i^t \right) \mid i = 1, \ldots, M,\ t = 0, \ldots, 23 \right\} \quad (6)$$

where each record corresponds to single draft turn in match m_i.

Identifiers:

- $match\_id \rightarrow m_i$ (match identifier)
- $user\_id \rightarrow$ corresponds to the team identity for drafting purposes.
- $timestamp \rightarrow$ draft turn index (0–23)

Action information:

- $action\_type \in \{\text{ban, pick}\} \rightarrow a^t$
- $hero\_action$ (e.g., hero_84_pick, hero_9_ban) $\rightarrow$ the action instance
- $item\_id$ interacted hero id

Outcome information:

- $label(0/1) \rightarrow$ indicates if the pick/ban as in Eq. (7).

$$label = \begin{cases} 1, & pick \\ 0, & ban \end{cases} \quad (7)$$

- $team\_result\ (0/1) \rightarrow$ team outcome for the corresponding side as in Eq.(8).

$$\text{team\_result} = \begin{cases} 1, & win \\ 0, & lose \end{cases} \quad (8)$$

- $patch \rightarrow$ game version, allows patch-wise analysis of drafts

Thus, each row encodes a tuple as in Eq. (9).

$$\left(s_i^t, a^t, c_i^t, y_i^t \right) \quad (9)$$

where:

- s_i^t (state) = ordered history of previous picks/bans + side/team metadata,
- a^t (action type) = ban or pick,
- c_i^t (hero) = the ground-truth hero banned or picked,

- y_i^t (label/outcome) = result indicator (win/loss context).

The sequence of rows with the same match_id yields the complete draft as in Eq. (10),

$$D_i = \left\{ \left(a^t, c_i^t \right) \right\}_{t=0}^{23} \tag{10}$$

which represents the real-world drafting process in esports.

Examining cross-domain distributions of users and items, Figure 3 plots users (x-axis) against items (y-axis) across categories, revealing strong differences in scale and density between domains. Building on this, Figure 4 reports the distribution of user–item interactions per category, highlighting how some domains are heavily imbalanced while others remain sparse. Structural variability is further emphasized in Figures 5 and 6, which show boxplots of shape and space per category, respectively. These results demonstrate that different domains exhibit highly divergent statistical properties, which complicates direct comparison and benchmarking.

DOTA-Draft was constructed to align with standard dataset properties while addressing these domain-specific gaps.

4.3. Preprocessing Pipeline

We developed our dataset creation pipeline to conform to established standards while bridging domain-specific gaps in in-game recommendation. As shown in Figure 7, the workflow begins with Dota 2 match data and proceeds through preprocessing, feature encoding, and export in a RecBole-compatible format. To ensure reproducibility and research suitability, the pipeline incorporates sensitivity analyses and baseline model training. The data are transformed into a research-ready benchmark through the following steps:

- De-duplication: Redundant entries were removed to ensure each match was uniquely represented.
- Entity normalization: Hero and team identifiers were cast to integers, and a composite team key (match_team_id) was constructed for consistency.
- Sessionization: Grouped draft actions per team per match and ordered them chronologically by draft sequence.
- Feature enrichment: Contextual attributes, including patch number, team outcome, hero name, action type (pick or ban), and draft order, were added.
- Label construction: Binary labels were assigned, with 1.0 for picks and 0.0 for bans.
- ID remapping: User and item identifiers were converted into sequential integer IDs, as required by RecBole.
- Schema definition: Structured with user, item, timestamp, label, and context fields for RecBole compatibility.
- Final dataset: The dataset is exported in standardized inter format for RecBole experiments.

5. Benchmarking & Evaluation

To better understand the properties of the proposed datasets, we analyze their core characteristics and present exploratory statistics. These analyses highlight the uniqueness of in-game recommendation data compared to traditional domains such as movies or e-commerce.

5.1 DOTA-Draft Characteristics

Following dataset construction, we conduct a detailed analysis of the temporal, structural, and strategic properties of DOTA-Draft to assess its suitability as a benchmark for in-game recommendation. The distribution of professional matches across patches exhibits clear segmentation patterns (Figure 8), reflecting the impact of balance updates and gameplay adjustments over time. This patch-wise separation enables longitudinal analysis and highlights the non-stationary nature of drafting data—an important departure from static recommendation environments.

A comparison between hero pick rates and win rates reveals pronounced discrepancies (Figure 9). Heroes that are frequently selected are not necessarily the most successful, indicating that popularity-driven signals alone are insufficient for optimal recommendation. This divergence introduces a critical modeling challenge: recommendation systems must distinguish between collective preference trends and competitive effectiveness, rather than relying solely on interaction frequency.

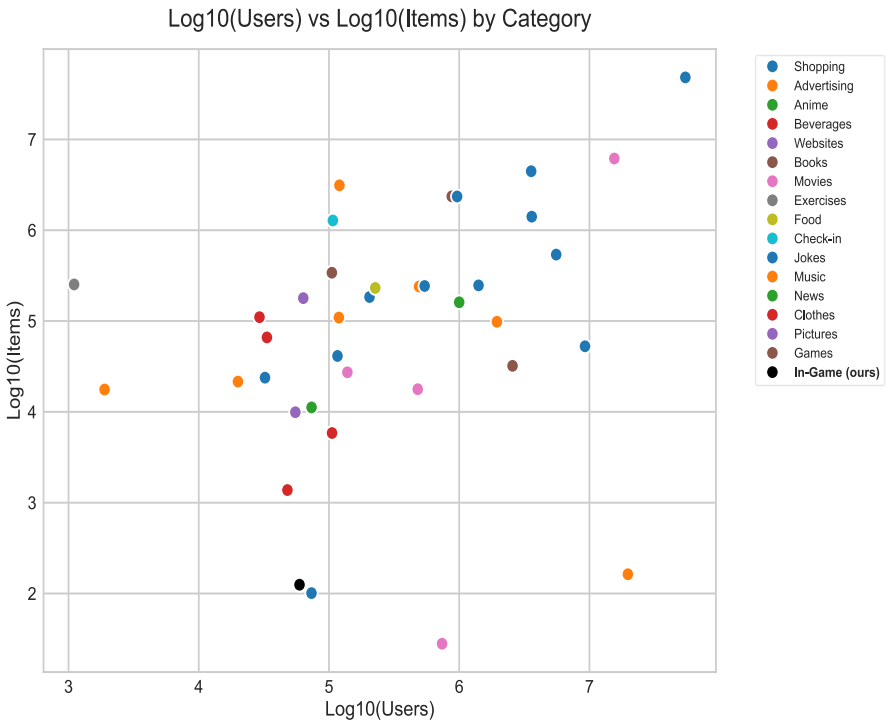

Figure 3. Plot of users (x-axis) and items (y-axis) across categories

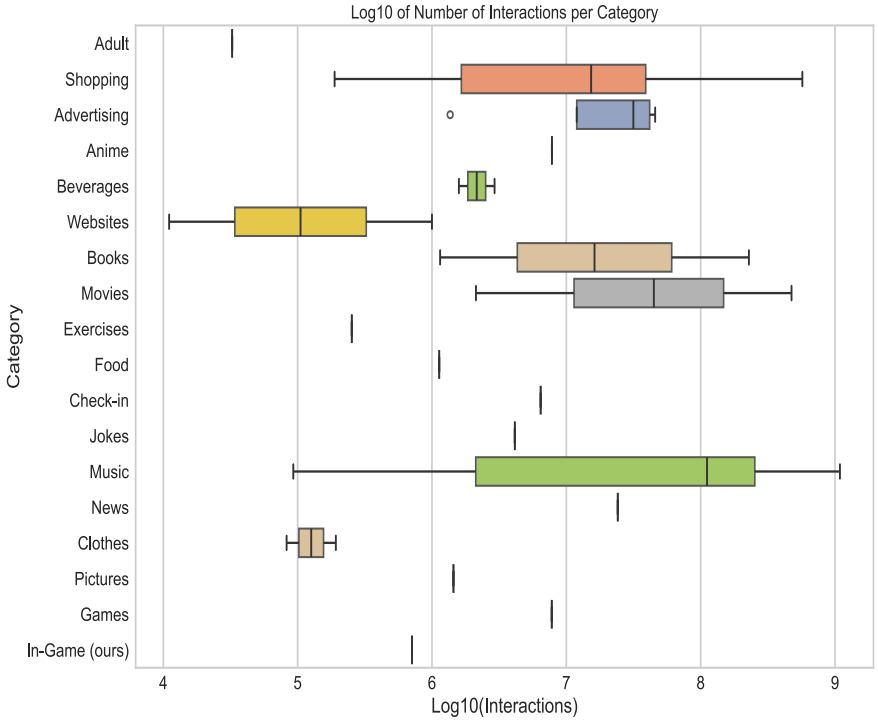

Figure 4. Boxplot of interactions per category.

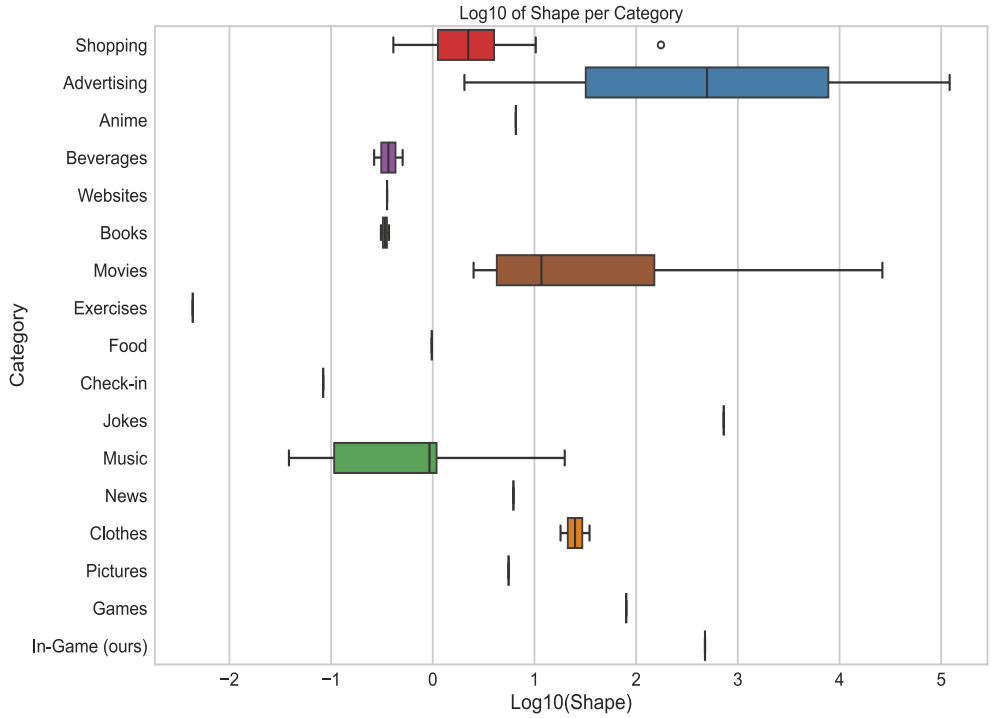

Figure 5. Boxplot of Shape per category.

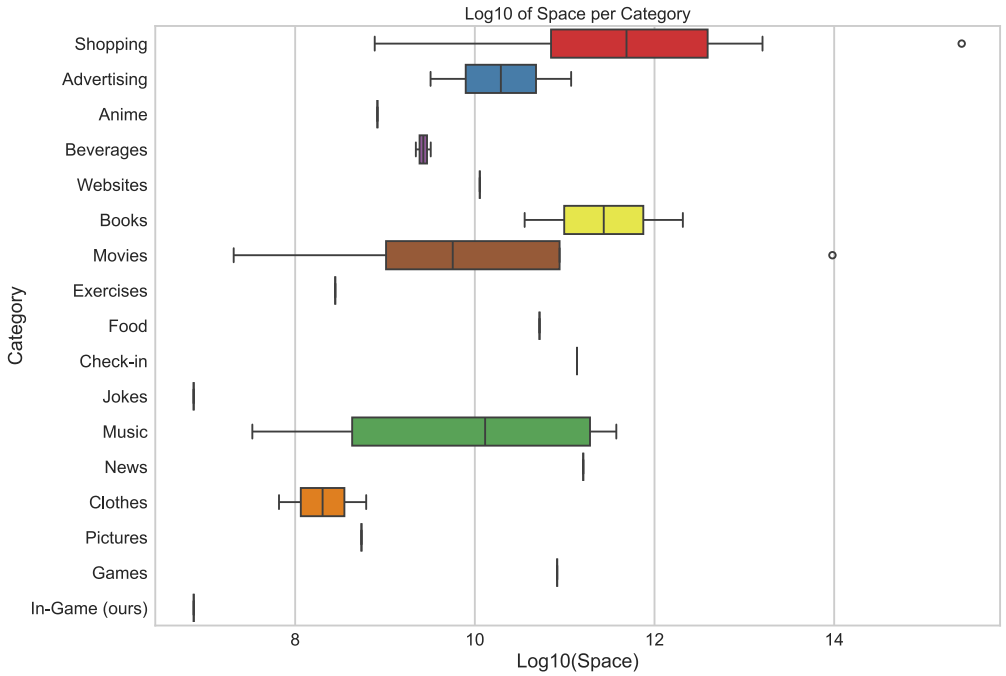

Figure 6. Boxplot of space per category.

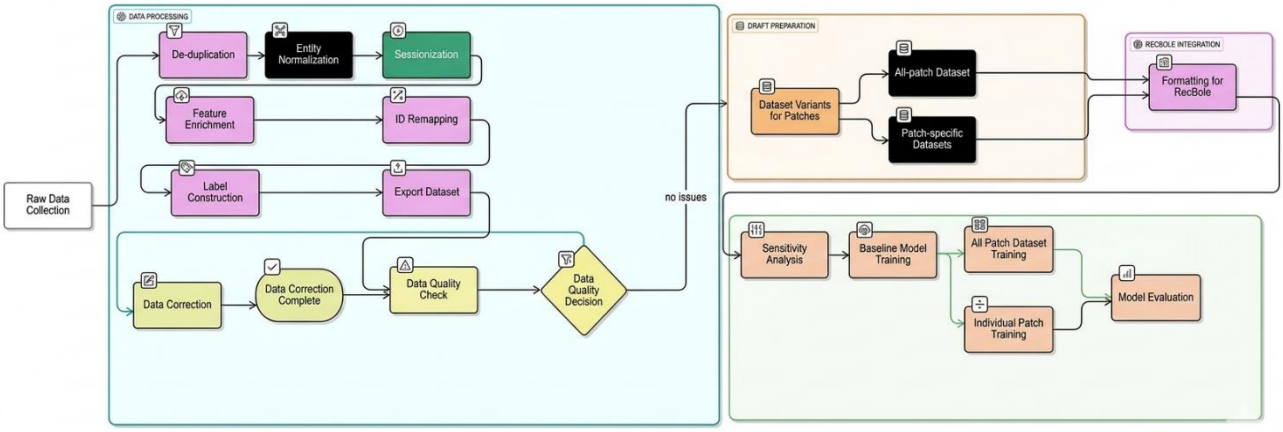

Figure 7. Workflow of dataset creation and analysis, from raw data preprocessing to RecBole export, sensitivity analyses, baseline model training, and evaluation.

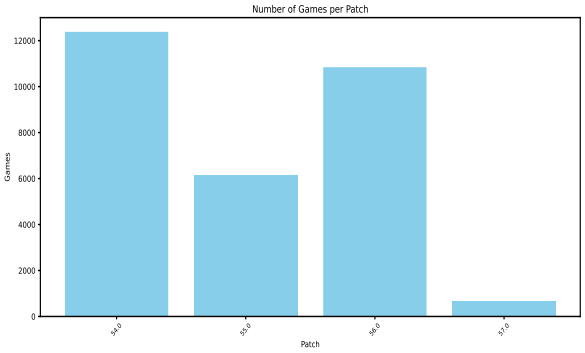

Figure 8. Number of games per patch, showing the distribution of matches across different patches.

A comparison between hero pick rates and win rates reveals pronounced discrepancies (Figure 9). Heroes that are frequently selected are not necessarily the most successful, indicating that popularity-driven signals alone are insufficient for optimal recommendation. This divergence introduces a critical modeling challenge: recommendation systems must distinguish between collective preference trends and competitive effectiveness, rather than relying solely on interaction frequency.

At the team level, win-rate variations between Radiant and Dire sides shift systematically across patches (Figure 10), suggesting that drafting outcomes are influenced by evolving structural asymmetries. These dynamics imply that recommendation models must account for contextual factors, including side information and temporal segmentation, to maintain robustness across balance changes.

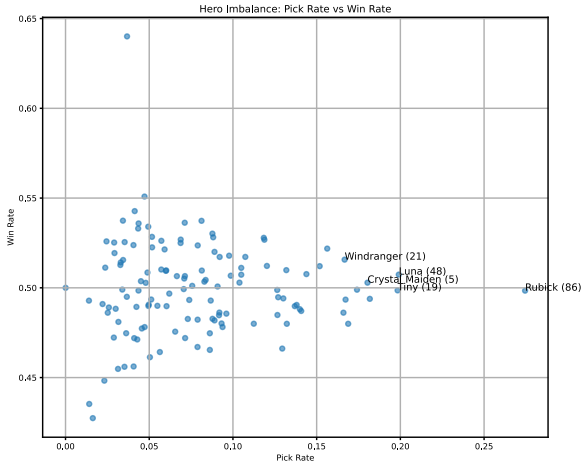

Figure 9. Hero pick rate versus win rate scatter plot. Heroes with extreme pick or win rates are labeled, providing a visual representation of performance imbalance.

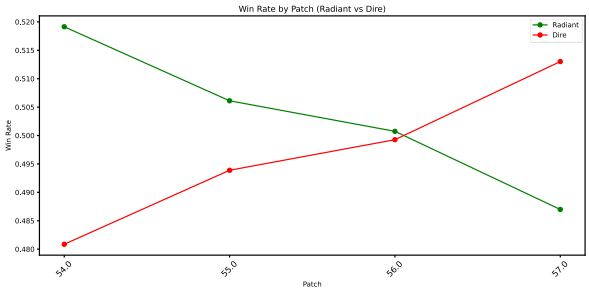

Figure 10. Win rate per patch for Dire and Radiant teams, showing performance differences between the two sides across patches.

Moreover, hero popularity—measured through combined pick and ban frequencies—demonstrates strong temporal fluctuation (Figure 11; see Appendix A, Figure A1 for the complete

distribution), reflecting meta-game evolution and adaptive strategic behavior. Because bans directly constrain the available action space, the drafting process constitutes a sequential, state-dependent decision problem rather than a static user–item interaction scenario. Each action modifies the feasible candidate set for subsequent decisions, introducing combinatorial dependencies that traditional collaborative filtering models are not designed to handle.

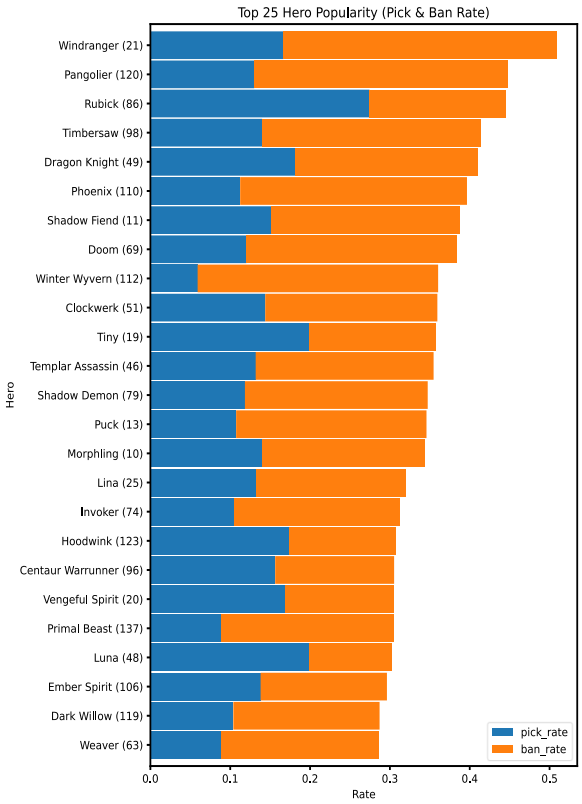

Figure 11. Horizontal bar chart illustrating the top 25 heroes by combined pick and ban frequency based on the DOTA-Draft dataset constructed in this study. Pick and ban frequencies are obtained by aggregating all drafting actions, where events are filtered by action type ("pick" or "ban") and counted for each hero. Each bar corresponds to a unique hero (identified by hero name and ID in the legend), with segments representing pick and ban counts. The visualization was generated by the authors from the processed dataset and does not reproduce or adapt previously published material.

Collectively, these characteristics show that DOTA-Draft captures both macro-level meta shifts and micro-level interaction dependencies. This dual complexity differentiates drafting data from conventional recommendation datasets, where items are typically independent and static. By preserving the full 24-step competitive drafting structure (12 bans and 12 picks), the dataset provides a realistic foundation for evaluating next-pick and top-k sequential recommendation under dynamic, multi-agent conditions.

The dataset comprises 29,327 professional matches, totaling 703,848 draft interactions across four competitive patches and 125 unique heroes. Its scale, temporal segmentation, and structured sequential encoding collectively establish DOTA-Draft as a principled benchmark for studying recommendation in adversarial, evolving environments.

5.2. Sensitivity Analysis

To lower the barrier for benchmarking, we report baseline results using BPR and GRU4Rec. Key hyperparameters, including learning rate and embedding size, were varied across predefined ranges—[0.0005, 0.001, 0.003, 0.01] for learning rate and [64, 96, 128] for embedding size—while other settings were held constant. Evaluation followed the protocol in Section 5, using Recall, NDCG, Hit Rate, and MRR to measure ranking performance. Tables 2 and 3 summarize quantitative outcomes for models trained on datasets segmented by patch. Sensitivity analyses reveal substantial variability in performance across hyperparameter configurations, highlighting the challenges of achieving stable recommendations in evolving game environments. To improve efficiency and reproducibility, hyperparameter searches were conducted on a 10% random subset of the training data, establishing DOTA-Draft as a principled benchmark for in-game recommendation research.

The results show that patch-specific evaluation provides higher performance compared to using the aggregated dataset. This indicates that the dynamics of the game environment—particularly changes introduced in each patch—have a strong influence on recommendation outcomes. In the context of Dota 2, patches often adjust hero abilities, item statistics, and overall balance. Such modifications alter player behavior and shift the distribution of hero selections, which directly impacts the recommender's effectiveness.

From a system design perspective, this emphasizes the importance of incorporating patch awareness or temporal modeling into recommendation approaches.

Ignoring patch-specific variations risks blending inconsistent interaction patterns, while leveraging patch information allows the model to better adapt to the evolving meta-game and improve predictive accuracy.

Table 2. Evaluation metrics (Recall@5, NDCG@5, Hit@5, MRR@5) for BPR and GRU4Rec models trained on different datasets: each patch separately and all patches combined.

Model	Recall@5					NDCG@5					Hit@5					MRR@5				
Patch	54	55	56	57	All	54	55	56	57	all	54	55	56	57	all	54	55	56	57	all
BPR	0.0897	0.118	0.1251	0.1345	0.0881	0.054	0.0752	0.0778	0.0833	0.054	0.0897	0.118	0.1251	0.1345	0.0881	0.0423	0.0613	0.0623	0.0666	0.043
GRU4Rec	0.2448	0.2471	0.2616	0.2621	0.2439	0.1596	0.157	0.1673	0.171	0.1614	0.2448	0.2471	0.2616	0.2621	0.2439	0.1317	0.1272	0.1364	0.1408	0.1343

Table 3. Evaluation metrics (Recall@10, NDCG@10, Hit@10, MRR@10) for BPR and GRU4Rec models trained on different datasets: each patch separately and all patches combined.

Model	Recall@10					NDCG@10					Hit@10					MRR@10				
Patch	54	55	56	57	All	54	55	56	57	all	54	55	56	57	all	54	55	56	57	all
BPR	0.1643	0.2028	0.2175	0.2341	0.1592	0.0778	0.1023	0.1073	0.1153	0.0768	0.1643	0.2028	0.2175	0.2341	0.1592	0.0521	0.0722	0.0743	0.0797	0.0522
GRU4Rec	0.3822	0.3623	0.4043	0.381	0.3749	0.2039	0.194	0.2132	0.2091	0.2035	0.3822	0.3623	0.4043	0.381	0.3749	0.1499	0.1423	0.1552	0.1563	0.1515

6. Discussion

The proposed benchmark dataset provides a structured foundation for advancing research in in-game recommendation. By systematically collecting, cleaning, and organizing interaction logs, the dataset bridges a crucial gap between existing general-purpose recommendation datasets (e.g., MovieLens, Amazon, Steam) and the unique dynamics of gaming environments. Unlike traditional datasets, where items are static and user preferences are often explicit, games introduce evolving states, time-sensitive interactions, and strategy-driven behaviors that pose distinctive challenges for recommender systems.

One key contribution of this work lies in demonstrating that recommendation in games cannot be effectively studied using static interaction data alone. The exploratory analysis highlights how user behavior in games is shaped by temporal factors (e.g., patch versions, balance updates), social contexts (e.g., team composition, cooperative play), and item characteristics (e.g., heroes, skills, or in-game assets). These dimensions demand benchmark datasets that explicitly encode game mechanics and context, enabling the development of models that go beyond conventional collaborative filtering. Another important consideration is the balance between dataset complexity and usability.

While granular gameplay logs offer rich contextual information, they can overwhelm researchers unfamiliar with specific game mechanics. To address this, the dataset is curated using standardized formats, accompanied by clear documentation and reproducible baselines. This approach lowers the entry barrier for broader adoption while retaining sufficient detail to support meaningful experimentation. Recommendation under non-stationary user behavior has also been addressed outside gaming, for instance by online models that adapt to shifts in user preference over time [62]; DOTA-Draft exposes a sharper form of this non-stationarity, driven by discrete patch releases rather than gradual drift.

The dataset also enables cross-disciplinary research beyond recommender systems. Its structure and richness support investigations into game balance, fairness in matchmaking algorithms, and the sociological dynamics of online gaming communities. This versatility underscores the broader impact of establishing a well-designed, accessible dataset for both technical and social research communities.

Finally, it is important to acknowledge the limitations of this work. While the dataset is comprehensive, its scope is inherently tied to specific game titles and contexts, which may limit generalizability across other genres or platforms without further adaptation. Additionally, privacy and ethical considerations remain paramount—ensuring proper anonymization, preventing harmful player profiling, and maintaining compliance with data protection regulations. Addressing these concerns will be essential for scaling the approach to broader gaming ecosystems.

7. Conclusion & Future Work

This work presents a benchmark dataset tailored for in-game recommendation, addressing a persistent gap within the recommender systems research community. Through systematic collection and curation of game interaction data, the dataset offers a structured resource that reflects the nuanced dynamics of gameplay—capturing temporal evolution, strategic choices, and contextual dependencies. Exploratory analyses reveal distinct player behavior patterns, underscoring the need for domain-specific

benchmarks that go beyond the assumptions of traditional datasets.

The proposed dataset not only establishes a foundation for evaluating recommendation algorithms in gaming contexts but also enables broader research directions, such as game balance and adaptive content personalization. Furthermore, by integrating our data-driven approach with existing advancements in procedural content generation (PCG), researchers can explore the synergy between behavioral personalization and automated content creation [63], fostering more dynamic and immersive gaming environments. This work effectively lowers the barrier for reproducible research while encouraging essential cross-disciplinary collaboration.

Looking ahead, several directions can further extend the contributions of this study. Future work may include:

- Expanding Dataset Coverage: Extending the dataset to multiple genres, platforms, and game modes to enhance generalizability and applicability.
- Dynamic Benchmarks: Incorporating evolving game patches, item rebalances, and meta-shifts to better reflect real-world conditions in competitive games.
- Evaluation Protocols: Designing task-specific evaluation metrics tailored for in-game recommendations, such as synergy prediction, adaptive matchmaking, or session-based next-action prediction.

In conclusion, this study contributes to establishing a robust research ecosystem for in-game recommendation. By addressing the long-standing absence of a dedicated benchmark dataset, it lays a foundation for advancing algorithmic innovation, standardized evaluation, and practical deployment. The resulting resource not only supports reproducible experimentation but also paves the way for the next generation of personalized gaming experiences.

References

[1] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context, " *Acm Trans. Interact. Intell. Syst.*, vol. 5, no. 4, pp. 1–19, 2015, doi: 10.1145/2827872.

[2] Y. Hou, J. Li, Z. He, A. Yan, X. Chen, and J. McAuley, "Bridging Language and Items for Retrieval and Recommendation, " no. 1, 2024, [Online]. Available: http://arxiv.org/abs/2403.03952

[3] I. Stewart, "Cups and downs, " *Coll. Math. J.*, vol. 43, no. 1, pp. 15–19, 2012, doi: 10.4169/college.math.j.43.1.015.

[4] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects, " in *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, 2019, pp. 188–197.

[5] E. Gibson, M. D. Griffiths, F. Calado, and A. Harris, "The relationship between videogame micro-transactions and problem gaming and gambling: A systematic review, " *Comput. Human Behav.*, vol. 131, p. 107219, 2022.

[6] S. Kühn, D. T. Kugler, K. Schmalen, M. Weichenberger, C. Witt, and J. Gallinat, "Does playing violent video games cause aggression? A longitudinal intervention study, " *Mol. Psychiatry*, vol. 24, no. 8, pp. 1220–1234, 2019.

[7] I. Granic, A. Lobel, and R. C. M. E. Engels, "The benefits of playing video games, " *Am. Psychol.*, vol. 69, no. 1, pp. 66–78, 2014, doi: 10.1037/a0034857.

[8] M. Quwaider, A. Alabed, and R. Duwairi, "The impact of video games on the players behaviors: A survey, " in *Procedia Computer Science*, Elsevier, 2019, pp. 575–582. doi: 10.1016/j.procs.2019.04.077.

[9] D. Forni, "Horizon Zero Dawn: The Educational Influence of Video Games in Counteracting Gender Stereotypes, " *Trans. Digit. Games Res. Assoc.*, vol. 5, no. 1, 2020, doi: 10.26503/todigra.v5i1.111.

[10] D. Madden, X. Liu, H. Yu, M. F. Sonbudak, G. M. Troiano, and C. Harteveld, "Why are you playing games? you are a girl!: Exploring gender biases in esports, " in *Conference on Human Factors in Computing Systems - Proceedings*, 2021, pp. 1–15. doi: 10.1145/3411764.3445248.

[11] L. Pascarella, F. Palomba, M. Di Penta, and A. Bacchelli, "How is video game development different from software development in open source?, " in *Proceedings - International Conference on Software Engineering*, 2018, pp. 392–402. doi: 10.1145/3196398.3196418.

[12] V. Nardone, B. Muse, M. Abidi, F. Khomh, and M. Di Penta, "Video Game Bad Smells: What They Are and How Developers Perceive Them, " *ACM Trans. Softw. Eng. Methodol.*, vol. 32, no. 4, pp. 1–35, 2023, doi: 10.1145/3563214.

[13] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE, " *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp. 109–122, 2020, doi: 10.21609/jiki.v13i2.885.

[14] J. J. Thompson, B. H. Leung, M. R. Blair, and M. Taboada, "Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model, " *Knowledge-Based Syst.*, vol. 137, pp. 149–162, 2017, doi: 10.1016/j.knosys.2017.09.022.

[15] A. Bucchiarone, K. M. L. Cooper, D. Lin, E. F. Melcer, and K. Sung, "Games and Software Engineering, " *ACM SIGSOFT Softw. Eng. Notes*, vol. 48, no. 1, pp. 85–89, 2023, doi: 10.1145/3573074.3573096.

[16] A. D'Angelo, C. Di Sipio, C. Politowski, and R. Rubei, *PlayMyData: a curated dataset of multi-platform video games*, vol. 1, no. 1. Association for Computing Machinery, 2024. doi: 10.1145/3643991.3644869.

[17] V. Araujo, H. Salinas, A. Labarca, A. Villa, and D. Parra, "Hierarchical Transformers for Group-Aware Sequential Recommendation: Application in MOBA Games, " *UMAP2022 - Adjun. Proc. 30th ACM Conf. User Model. Adapt. Pers.*, pp. 293–301, 2022, doi: 10.1145/3511047.3537667.

[18] L. Hanke and L. Chaimowicz, "A recommender system for hero line-ups in MOBA games, " *Proc. 13th AAAI Conf. Artif. Intell. Interact. Digit. Entertain. AIIDE 2017*, pp. 43–49, 2017, doi: 10.1609/aiide.v13i1.12938.

[19] M. Viggiato and C. P. Bezemer, "Trouncing in dota 2: An investigation of blowout matches, " *Proc. 16th AAAI Conf. Artif. Intell. Interact. Digit. Entertain. AIIDE 2020*, pp. 294–300, 2020, doi: 10.1609/aiide.v16i1.7444.

[20] J. M. Wang-Cheng Kang, "Self-Attentive Sequential Recommendation, " *IEEE Int. Conf. Data Min.*, vol. 2018-Novem, pp. 197–206, 2018.

[21] M. Wan and J. McAuley, "Item recommendation on monotonic behavior chains, " *RecSys 2018 - 12th ACM Conf. Recomm. Syst.*, pp. 86–94, 2018, doi: 10.1145/3240323.3240369.

[22] A. Pathak, K. Gupta, and J. McAuley, "Generating and personalizing bundle recommendations on steam, " *SIGIR 2017 - Proc. 40th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 1073–1076, 2017, doi: 10.1145/3077136.3080724.

[23] BwandoWando, "Dota 2 Pro League Matches 2016-2025. " Accessed: Apr. 20, 2026. [Online]. Available: https://www.kaggle.com/datasets/bwandowando/dota-2-pro-league-matches-2023/

[24] K. Akhmedov and A. H. Phan, "Machine learning models for DOTA 2 outcomes prediction, " *arXiv Prepr. arXiv2106.01782*, 2021.

[25] K. Conley and D. Perry, "How Does He Saw Me? A Recommendation Engine for Picking Heroes in Dota 2, " *CS229 Previous Proj.*, vol. 7, 2013, [Online]. Available: https://cs229.stanford.edu/proj2013/PerryConley-HowDoesHeSawMeARecommendationEngineForPickingHeroesInDota2.pdf

[26] A. Drachen, M. Yancey, J. Maguire, D. Chu, I. Y. Wang, T. Mahlmann, M. Schubert, and D. Klabajan, "Skill-based differences in spatio-Temporal team behaviour in defence of the Ancients 2 (DotA 2), " in *Conference Proceedings - 2014 IEEE Games, Media, Entertainment Conference, IEEE GEM 2014*, 2015, pp. 1–8. doi: 10.1109/GEM.2014.7048109.

[27] T. E. Batsford, "Calculating Optimal Jungling Routes in DOTA2 Using Neural Networks and Genetic Algorithms, " *Game Behav.*, vol. 1, no. 1, 2014.

[28] T. Flint, "Creating a Hero Recommender System for Newer Player in Dota 2, " 2022.

[29] H. Lee, D. Hwang, H. Kim, B. Lee, and J. Choo, "DraftRec: Personalized Draft Recommendation for Winning in Multi-Player Online Battle Arena Games, " in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, Association for Computing Machinery, 2022, pp. 3428–3439. doi: 10.1145/3485447.3512278.

[30] A. Summerville, M. Cook, and B. Steenhuisen, "Draft-Analysis of the ancients: Predicting draft picks in DotA 2 using machine learning, " *AAAI Work. - Tech. Rep.*, vol. WS-16-21-WS-16-23, no. Godec, pp. 100–106, 2016.

[31] Z. Chen, T.-H. D. Nguyen, Y. Xu, C. Amato, S. Cooper, Y. Sun, and M. S. El-Nasr, "The Art of Drafting: A Team-Oriented Hero Recommendation System for Multiplayer Online Battle Arena Games, " 2018, [Online]. Available: http://arxiv.org/abs/1806.10130

[32] W. X. Zhao, S. Mu, Y. Hou, Z. Lin, Y. Chen, X. Pan, K. Li, Y. Lu, H. Wang, C. Tian, Y. Min, Z. Feng, X. Fan, X. Chen, P. Wang, W. Ji, Y. Li, X. Wang, and J. R. Wen, "RecBole: Towards a Unified, Comprehensive and Efficient Framework for Recommendation Algorithms, " *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 4653–4664, 2021, doi: 10.1145/3459637.3482016.

[33] W. X. Zhao, Y. Hou, X. Pan, C. Yang, Z. Zhang, Z. Lin, J. Zhang, S. Bian, J. Tang, W. Sun, Y. Chen, L. Xu, G. Zhang, Z. Tian, C. Tian, S. Mu, X. Fan, X. Chen, and J. R. Wen, *RecBole 2.0: Towards a More Up-to-Date Recommendation Library*, vol. 1, no. 1. Association for Computing Machinery, 2022. doi: 10.1145/3511808.3557680.

[34] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback, " *Proc. 25th Conf. Uncertain. Artif. Intell. UAI 2009*, pp. 452–461, 2009.

[35] F. A. N. Ge, C. Zhang, K. Wang, L. I. Yingjie, J. Chen, and X. U. Zenglin, "CUPID: Improving Battle Fairness and Position Satisfaction in Online MOBA Games with a Re-matchmaking System, " *Proc. ACM Human-Computer Interact.*, vol. 8, no. CSCW2, 2024,

doi: 10.1145/3686978.

[36] K. W. Błaszczyk and D. Szajerman, "Champion Recommendation in League of Legends Using Machine Learning, " *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14076 LNCS, pp. 155–170, 2023, doi: 10.1007/978-3-031-36027-5_12.

[37] C. Chen, F. Mo, X. Fan, C. Bai, and H. Yamana, "Mobarec-gcnfp: Champion recommendation for multi-player online battle arena games using graph convolution network with fewer parameters, " in *2023 IEEE 8th International Conference on Big Data Analytics (ICBDA)*, 2023, pp. 147–153.

[38] J. Y. Chin, Y. Chen, and G. Cong, "The datasets dilemma: How much DoWe really know about recommendation datasets?, " *WSDM 2022 - Proc. 15th ACM Int. Conf. Web Search Data Min.*, no. 3, pp. 141–149, 2022, doi: 10.1145/3488560.3498519.

[39] B. Becker and R. Kohavi, "Adult, " 1996, doi: 10.24432/C5XW20.

[40] J.-B. Tien, joycenv, and O. Chapelle, "Display Advertising Challenge. " Accessed: Apr. 21, 2026. [Online]. Available: https://kaggle.com/competitions/criteo-display-ad-challenge.

[41] S. Wang and W. Cukierski, "Click-Through Rate Prediction. " Accessed: Apr. 21, 2026. [Online]. Available: https://kaggle.com/competitions/avazu-ctr-prediction

[42] W. Zhang, S. Yuan, J. Wang, and X. Shen, "Real-Time Bidding Benchmarking with iPinYou Dataset, " 2014, [Online]. Available: http://arxiv.org/abs/1407.7073

[43] J. McAuley and J. Leskovec, "From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews, " *WWW 2013 - Proc. 22nd Int. Conf. World Wide Web*, pp. 897–907, 2013.

[44] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks, " in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 1082–1090.

[45] R. Misra, M. Wan, and J. McAuley, "Decomposing fit semantics for product size recommendation in metric spaces, " *RecSys 2018 - 12th ACM Conf. Recomm. Syst.*, pp. 422–426, 2018, doi: 10.1145/3240323.3240398.

[46] J. Ni, L. Muhlstein, and J. McAuley, "Modeling heart rate and activity data for personalized fitness recommendation, " *Web Conf. 2019 - Proc. World Wide Web Conf. WWW 2019*, vol. 2, pp. 1343–1353, 2019, doi: 10.1145/3308558.3313643.

[47] J. Stamper, A. Niculescu-Mizil, S. Ritter, G. J. Gordon, and K. R. Koedinger, "Algebra I 2008-2009: Challenge data set from KDD Cup 2010 Educational Data Mining Challenge, " 2010. [Online]. Available: http://pslcdatashop.web.cmu.edu/KDDCup/downloads.jsp

[48] B. P. Majumder, S. Li, J. Ni, and J. McAuley, "Generating personalized recipes from historical user preferences, " *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 5976–5982, 2019, doi: 10.18653/v1/d19-1613.

[49] F. Zhu, Y. Wang, C. Chen, G. Liu, and X. Zheng, "A graphical and attentional framework for dual-target cross-domain recommendation, " *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2021-Janua, pp. 3001–3008, 2020, doi: 10.24963/ijcai.2020/415.

[50] J. Rappaz, J. McAuley, and K. Aberer, "Recommendation on live-streaming platforms: Dynamic availability and repeat consumption, " in *RecSys 2021 - 15th ACM Conference on Recommender Systems*, New York, NY, USA: ACM, Sep. 2021, pp. 390–399. doi: 10.1145/3460231.3474267.

[51] S. Oramas, V. C. Ostuni, T. Di Noia, X. Serra, and E. Di Sciascio, "Sound and music recommendation with knowledge graphs, " *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 2, pp. 21:1--21:21, 2016, doi: 10.1145/2926718.

[52] I. Cantador, P. Brusilovsky, and T. Kuflik, "Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011), " in *Proceedings of the fifth ACM conference on Recommender systems*, in RecSys 2011. New York, NY, USA: ACM, Oct. 2011, pp. 387–388. doi: 10.1145/2043932.2044016.

[53] M. Schedl, "The LFM-1b Dataset for Music Retrieval and Recommendation, " in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, New York, NY, USA: ACM, Jun. 2016, pp. 103–110. doi: 10.1145/2911996.2912004.

[54] M. Moscati, E. Parada-Cabaleiro, Y. Deldjoo, E. Zangerle, and M. Schedl, "Music4All-Onion - A Large-Scale Multi-faceted Content-Centric Music Recommendation Dataset, " in *International Conference on Information and Knowledge Management, Proceedings*, M. Al Hasan and L. Xiong, Eds., New York, NY, USA: ACM, Oct. 2022, pp. 4339–4343. doi: 10.1145/3511808.3557656.

[55] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, and M. Zhou, "MIND: A Large-scale Dataset for News Recommendation, " in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 3597–3606. doi: 10.18653/v1/2020.acl-main.331.

[56] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural Collaborative Filtering, " in *Proceedings of the 26th International Conference on World Wide*

Web, Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, Apr. 2017, pp. 173–182. doi: 10.1145/3038912.3052569.

[57] D. P. KOHEN, K. N. OLNESS, S. O. COLWELL, and A. HEIMEL, "The Use of Relaxation-Mental Imagery (Self-Hypnosis) in the Management of 505 Pediatric Behavioral Encounters, " *J. Dev. Behav. Pediatr.*, vol. 5, no. 1, p. 21???25, Feb. 1984, doi: 10.1097/00004703-198402000-00005.

[58] Y. Zhang, M. Zhang, Y. Liu, S. Ma, and S. Feng, "Localized matrix factorization for recommendation based on matrix block diagonal forms, " in *Proceedings of the 22nd international conference on World Wide Web*, New York, NY, USA: ACM, May 2013, pp. 1511–1520. doi: 10.1145/2488388.2488520.

[59] R. Zykov, N. Artem, and A. Alexander, "RetailRocket Recommender System Dataset, " Kaggle. Accessed: Apr. 21, 2026. [Online]. Available: https://www.kaggle.com/datasets/retailrocket/ecommerce-dataset

[60] R. Mohammad and L. McCluskey, "Phishing Websites, " UCI Machine Learning Repository. Accessed: Apr. 21, 2026. [Online]. Available: https://archive.ics.uci.edu/dataset/327/phishing+websites

[61] R. He, C. Fang, Z. Wang, and J. McAuley, "Vista: A visually, socially, and temporally-aware model for artistic recommendation, " in *RecSys 2016 - Proceedings of the 10th ACM Conference on Recommender Systems*, 2016, pp. 309–316. doi: 10.1145/2959100.2959152.

[62] J. Hamidzadeh and M. Moradi, "Online recommender system considering changes in user's preference, " J. AI Data Min., vol. 9, no. 2, pp. 203–212, 2021, doi: 10.22044/jadm.2020.9518.2085.

[63] S. Saffari, M. Dorrigiv and F. Yaghmaee, "Harnessing Machine Learning for Procedural Content Generation in Gaming: A Comprehensive Review, " J. AI Data Min., vol. 12, no. 4, pp. 583–597, 2024, doi: 10.22044/jadm.2025.15016.2603.

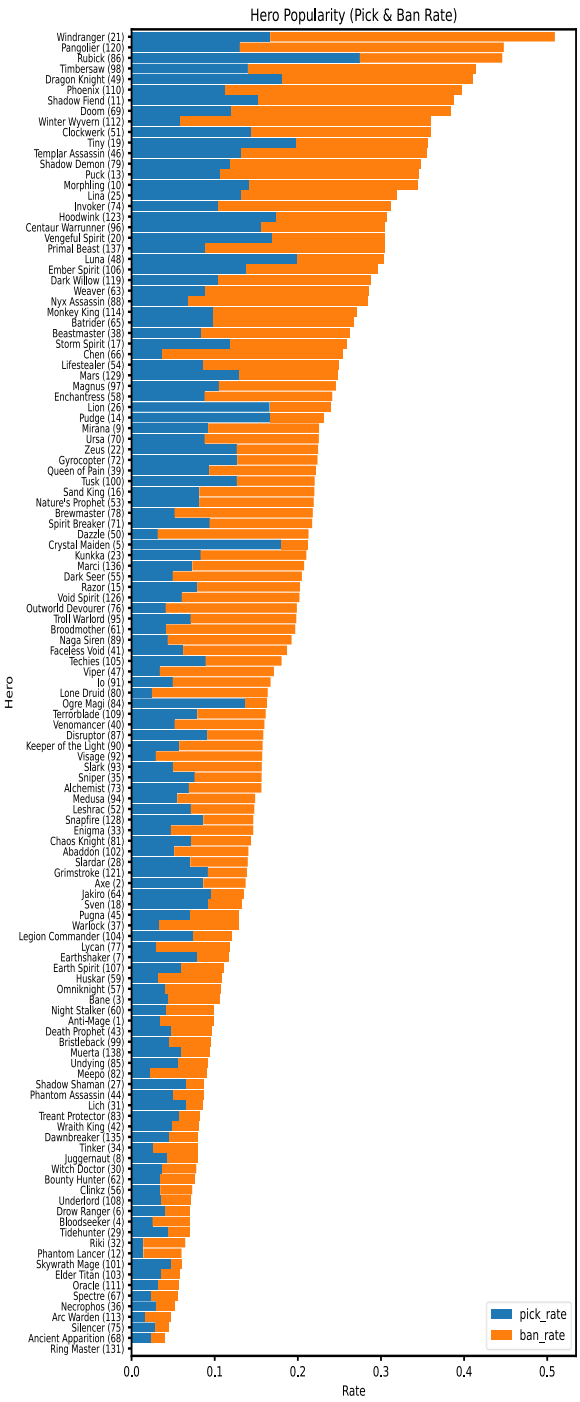

Figure A1. Horizontal bar chart showing complete hero popularity in the DOTA-Draft dataset constructed in this study. Pick and ban frequencies are computed by aggregating all drafting actions, filtering by action type ("pick" or "ban"), and counting occurrences for each hero. Each bar represents a unique hero (identified by name and ID in the legend), with segments indicating pick and ban counts. This appendix figure presents the full hero distribution, whereas Figure 11 in the main manuscript displays only the top 25 heroes for improved readability. This is an original visualization generated by the authors and does not reproduce or adapt any previously published material.

Appendix

Appendix A. Complete Hero Popularity Distribution

Figure A1 presents the complete hero popularity distribution derived from the DOTA-Draft dataset constructed in this study. In the main manuscript, Figure 11 was revised to display only the top 25 heroes with the highest combined pick and ban frequencies in order to improve readability and visualization clarity in the manuscript format. The complete distribution is provided here for completeness and reproducibility.

Hero popularity statistics were computed by aggregating drafting actions across all matches, where events were filtered according to action type ("pick" or "ban") and counted separately for each hero. Each horizontal bar corresponds to a unique hero identified by hero ID and hero name, while bar segments indicate pick and ban frequencies.

The visualization was generated by the authors from the processed dataset and is included as an original figure derived from this study.

ایجاد مجموعه‌دادگان برای سیستم‌های توصیه‌گر درون‌بازی: پژوهشی مجموعه‌دادگان‌محور

محمدرضا محمدنژاد، مرتضی دُرّی‌گیو* و فرزین یغمایی

دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران.

ارسال ۲۰۲۵/۱۰/۰۴؛ بازنگری ۲۰۲۵/۱۲/۱۹؛ پذیرش ۲۰۲۶/۰۵/۲۰

چکیده:

پژوهش‌ها در حوزه سیستم‌های توصیه‌گر عمدتاً بر مجموعه‌دادگان استاندارد نظیر MovieLens، Amazon Reviews و Last.fm متکی بوده‌اند. با این حال، به دلیل ماهیتِ توالی‌محور، تیمی و رقابتیِ روند بازی در بازی‌های «عرصه نبرد چندنفره برخط» (MOBA)، این مجموعه‌دادگان برای توصیه‌گری درون‌بازی مناسب نیستند. ما برای شناسایی ویژگی‌های ضروری در مجموعه‌دادگانِ توصیه‌گر درون‌بازی، یک تحلیلِ میان‌دامنه‌ای بر روی مجموعه‌دادگان پرکاربرد انجام داده و ویژگی‌های ساختاری و توزیعی آن‌ها از جمله فضای تعامل، شکل ماتریس، پراکندگی و تنوعِ ویژگی-شکل مبتنی بر ضریب جینی را ارزیابی کرده‌ایم. با بهره‌گیری از این یافته‌ها، ما «DOTA-Draft» را به عنوان یک مجموعه‌دادگانِ آماده‌سازی‌شده برای پژوهش، از داده‌های خام مسابقات حرفه‌ای Dota 2 گردآوری کرده‌ایم که وضعیتِ توالیِ انتخاب/ممنوعیت (pick/ban)، نسخه‌های پچ و نتایج مسابقات را کدگذاری می‌کند. با استفاده از این مجموعه‌دادگان، ما وظایفِ توصیه‌گریِ انتخابِ برتر (top-k) را انجام داده و نتایج مبنا (baseline) را با استفاده از الگوریتم‌های BPR و GRU4Rec ارائه کرده‌ایم. به‌منظور تسهیل در به‌کارگیری، DOTA-Draft در قالب سازگار با کتابخانه RecBole بسته‌بندی شده است. این پژوهش، معیارهای اصولی برای توصیه‌گری درون‌بازی تعریف کرده، ناکارآمدی پارادایم‌های سنتیِ کاربر-آیتم را در محیط‌های پویا و رقابتی اثبات می‌کند و زیربنایی برای توسعه مدل‌هایی فراهم می‌سازد که تصمیم‌گیری‌های توالی‌محور و چندعاملی را مد نظر قرار می‌دهند.

کلمات کلیدی: سیستم‌های توصیه‌گر درون‌بازی، عرصه نبرد چندنفره برخط (MOBA)، مجموعه‌دادگان تعاملی دوتا ۲ (Dota 2)، مدل‌های توصیه‌گر مبتنی بر نشست، ارزیابی معیار با چارچوب RecBole.