University of Antwerp
Faculty of Science

# BALANCING ACCURACY & DIVERSITY

## in Steam Game Recommendations

Final Presentation | Artificial Intelligence Project

**Hala Alramli**

# 1. RESEARCH QUESTIONS & RELEVANCE

## THE PROBLEM

**"Echo Chambers"** - Users get bored with redundant **recommendations** (e.g., 5 versions of the same shooter).

Traditional algorithms optimize for **Accuracy** only, ignoring **Novelty**.

> "What is the impact of applying a **diversity re-ranking algorithm (MMR)** on the **accuracy-diversity trade-off?**"

## HYPOTHESIS


We hypothesize that we can significantly **boost Intra-List Diversity (ILD)** with only a marginal, acceptable loss in **Ranking Accuracy (NDCG)**.

**Goal:** Find the **"Sweet Spot"** where users discover new genres without losing relevance.

# 2. METHODOLOGY & DEFENSE

### STEP 1: DATA SPLIT

**CHOICE**

Strong Generalization

**DEFENSE / RATIONALE**

Prevents *Data Leakage*. Simulates real-world scenario of completely new users.

### STEP 2: FILTERING

**CHOICE**

5-core

**DEFENSE / RATIONALE**

Ensures **model stability** by removing noise (users/items with too few interactions).

### STEP 3: FEATURES

**CHOICE**

Tags + Genres

**DEFENSE / RATIONALE**

Excluded 'Specs' as they acted as noise, artificially inflating similarity metrics.

### STEP 4: MODEL

**CHOICE**

EASE + MMR

**DEFENSE / RATIONALE**

EASE (Shallow Autoencoder) for high accuracy + MMR for diversity re-ranking.

# 3. RESULTS: BASELINE SELECTION

Comparing traditional neighbor-based methods vs. shallow autoencoders.


***Winner***: EASE outperformed UserKNN by **~3%** in **Accuracy** and **~3% in Diversity**, proving superior for sparse Steam data.

# 4. RESULTS: THE TRADE-OFF FRONTIER (EASE + MMR)

| | NDCG (Acc) | ILD (Div) |
|---|---|---|
| EASE ($\lambda$=1.0) | 0.3733 | 0.6254 |
| MMR ($\lambda$=0.7) | 0.3573 ↓ | 0.6970 ↑ |
| MMR ($\lambda$=0.5) | 0.3216 | 0.7370 |

## Λ = 0.7

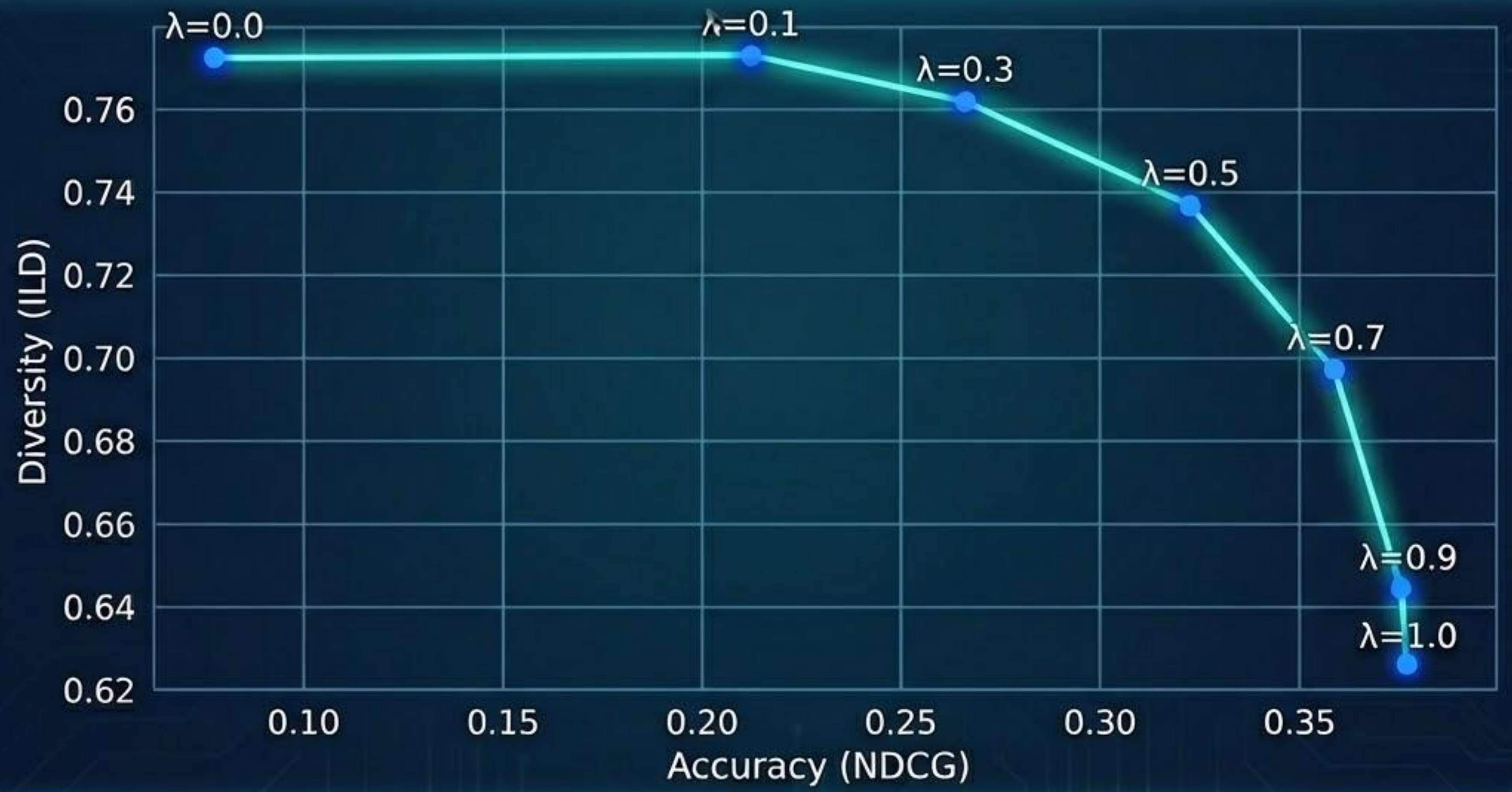
**OPTIMAL BALANCE**

**+7%** Diversity

**-2%** Accuracy

This confirms our hypothesis: a small sacrifice in theoretical accuracy yields a massive gain in content variety.

# Accuracy (NDCG) vs. Diversity (ILD) Trade-off @ K=20 (EASE + MMR)



# 5. EXTERNAL VALIDATION (CODEBENCH)

Performance on held-out test set validates the internal findings.


| Model | NDCG | ILS (Lower is better) | Novelty |
|---|---|---|---|
| EASE (Baseline) | 0.3888 | 0.2709 | 0.7572 |
| UserKNN | 0.3533 | 0.3468 | 0.7433 |
| MMR (λ=0.7) | 0.3653 ↓ | 0.2374 ↓ | 0.7774 ↑ |
| MMR (λ=0.5) | 0.3264 | 0.2126 ↓ | 0.7913 ↑ |

λ=0.5 maximizes Novelty for users who strictly prefer exploration over relevance.

# 6. DISCUSSION & CONCLUSION

## KEY FINDINGS

- **Feature Engineering:** Removing "Specs" metadata was crucial; it acted as noise in diversity calculations.
- **Pareto Optimality:** Diversity is not a zero-sum game. We can improve user experience without destroying relevance.

## LIMITATIONS

- **Offline Gap:** High ILD implies diversity, but only online A/B testing can prove actual user satisfaction.
- **Cold Start:** Users with < 5 items were filtered; they require a Content-based strategy.