AIE425 Intelligent Recommender Systems, Fall Semester 25/26

FINAL COURSE PROJECT PLAGRISM REPORT

GROUP 4

221101140   NOURELDEEN AHMED MAHER MESBAH

221101573   YOUSEF MOHAMED ABDELWHAB

221101030   YOUSEF ZAKARIA SOBHY

221100244   BASSANT KAMAL MESILAM

SUBMISSION DATE: MONDAY, JANUARY 5,2026

As we don't have access to Paid Plagiarism Checker Tools , we will be taking 5 random samples from the text inside the report and uploading each of them to 2 different Free tools for checking (Grammarly & ZeroGPT)

Plagiarism Checker

Ensure every word is your own with Grammarly's AI-powered plagiarism checker, which uses advanced AI to detect plagiarism in your text and check for other writing issues.

of point 4 (MLE Top 5)

Methods Compared:

- Mean-Fill Point 9: k-NN with cosine similarity in 5D latent space
- MLE Point 4: Reconstruction formula ($\mu_i + \Sigma\ t_{u,p} \times W_{i,p}$) in 5D latent space

Prediction Results (Top-5 PCs)

User	Item	Mean-Fill (Point 9)	MLE (Point 4)	Better
U1	I1	3.30 (err=0.39)	3.69 (err=0.00)	MLE

Scan for plagiarism   ↑ Upload a file   6154/10000

19 We didn't find any plagiarism, but we found 19 writing issues.

No plagiarism found ✓   Grammar 🔟

Spelling ❶   Punctuation ✓

Conciseness ✓   Readability ✓

Word choice ❹   Additional issues ❹

Get Grammarly Pro

Aspect	Winner
Prediction Accuracy	MLE
Variance Explained	MLE
Statistical Validity	MLE
Computational Simplicity	Mean-Fill
Overall	MLE

Maximum Likelihood Estimation fundamentally transforms PCA-based recommendation by respecting the observed data structure, leading to dramatically better rating predictions

Scan for plagiarism   ↑ Upload a file   5134/10000

15 We didn't find any plagiarism, but we found 15 writing issues.

No plagiarism found ✓   Grammar ❼

Spelling ✓   Punctuation ❶

Conciseness ✓   Readability ✓

Word choice ❸   Additional issues ❹

Get Grammarly Pro

Plagiarism Checker

Ensure every word is your own with Grammarly's AI-powered plagiarism checker, which uses advanced AI to detect plagiarism in your text and check for other writing issues.

2.3 Feature Extraction (Content-Based)

To understand streamer content, we extracted features from:

1. Text: Game titles, genres, and themes (from IGDB).

o   Technique: TF-IDF Vectorization (max_features=1000, min_df=2).

2. Numerical: Popularity metrics (Average Viewers, Followers).

o   Technique: Log-transformation (log1p) followed by Min-Max Scaling to handle power-law usage distribution.

3. Combination: Features were weighted (90% Text / 10% Numerical) to prioritize content relevance over raw popularity.

Scan for plagiarism   ↑ Upload a file   8409/**10000**

---

28   We didn't find any plagiarism, but we found 28 writing issues.

No plagiarism found	✓	Grammar	9
Spelling	3	Punctuation	8
Conciseness	✓	Readability	✓
Word choice	1	Additional issues	7

Get Grammarly Pro

optimized SVD), delivering statistically significant improvements over non-hybrid baselines

7. Appendices

Appendix A: Key Code Snippets

Snippets from

- Collaberative.py
- Hybrid.py
- Content_based.py

Scan for plagiarism   ↑ Upload a file   6494/**10000**

!   **We have found plagiarism** in your text and have also detected 17 writing issues.

Plagiarism found	!	Grammar	3
Spelling	2	Punctuation	✓
Conciseness	1	Readability	✓
Word choice	✓	Additional issues	11

Get Grammarly Pro

Plagiarism Checker

Ensure every word is your own with Grammarly's AI-powered plagiarism checker, which uses advanced AI to detect plagiarism in your text and check for other writing issues.

start users, and cascade hybrid architectures should be considered for balancing accuracy with computational efficiency.

Future Directions

Future work should explore deep learning approaches for latent factor extraction, dynamic user profile modeling to capture preference drift over time, and integration of additional contextual signals such as temporal patterns and social connections. The successful implementation of the Twitch.tv hybrid system provides a foundation for extending these techniques to other streaming and content recommendation domains.

Scan for plagiarism ↑ Upload a file   3148/**10000**

7 We didn't find any plagiarism, but we found 7 writing issues.

Plagiarism found	✓	Grammar	❷
Spelling	✓	Punctuation	✓
Conciseness	✓	Readability	✓
Word choice	✓	Additional issues	❺

Get Grammarly Pro

Trusted GPT-5, ChatGPT and AI Detector tool by ZeroGPT

ZeroGPT the most Advanced and Reliable Chat GPT, GPT5 & AI Content Detector

```
o      Top-5: Less computation, good for sparse users
o      Top-10: Better accuracy, recommended for active users

1.6-   Compare the results of point 9 in part 1 (Mean-fill Top 5) with results of point 4 (MLE Top 5)

Methods Compared:
•      Mean-Fill Point 9: k-NN with cosine similarity in 5D latent space
•      MLE Point 4: Reconstruction formula (μ_i + Σ t_u,p × W_i,p) in 5D latent space

Prediction Results (Top-5 PCs)
User   Item   Mean-Fill (Point 9)   MLE (Point 4)   Better
U1     I1     3.30 (err=0.39) 3.69 (err=0.00) MLE
U1     I2     3.10 (err=0.64) 3.74 (err=0.00) MLE
```

Detect Text   ↑ Upload File   6,154/15,000 Characters
Check 350,000 characters, **Upgrade Here**

Your Text is Most Likely Human written, may include parts generated by AI/GPT

**24.16%
AI GPT***

Analysis and Comments

1. Top-10 PCs Generally Perform Better

Top-10 PCs achieve lower average error (0.39) compared to Top-5 PCs (0.51). This is expected because:

Trusted GPT-5, ChatGPT and AI Detector tool by ZeroGPT

ZeroGPT the most Advanced and Reliable Chat GPT, GPT5 & AI Content Detector

PCA MLE with Top-10 Principal Components is the recommended approach for rating prediction.
The computational overhead of MLE is justified by its significantly superior prediction accuracy and
statistical validity.

Final Verdict
Aspect Winner
Prediction Accuracy   MLE
Variance Explained   MLE
Statistical Validity   MLE
Computational Simplicity   Mean-Fill
Overall MLE
Maximum Likelihood Estimation fundamentally transforms PCA-based recommendation by respecting the observed data
structure, leading to dramatically better rating predictions

Detect Text   ⬆ Upload File   5,134/15,000 Characters
Check 350,000 characters, **Upgrade Here**

Your Text is Most Likely Human written, may include parts generated by AI/GPT

**26.39%
AI GPT***

Prediction Results (Top-10 PCs)

User Item Mean-Fill (Point 11) MLE (Point 6) Better

U1 I1 3.55 (err=0.14) 3.69 (err=0.00) MLE

Trusted GPT-5, ChatGPT and AI Detector tool by ZeroGPT

ZeroGPT the most Advanced and Reliable Chat GPT, GPT5 & AI Content Detector

- Users: ~90,000 unique users.
- Items (Streamers): ~1,400 active streamers.
- Rating Scale: Implicit feedback converted to 1-5 scale (based on watch time/frequency).

2.3 Feature Extraction (Content-Based)
To understand streamer content, we extracted features from:
1. Text: Game titles, genres, and themes (from IGDB).
o Technique: TF-IDF Vectorization (max_features=1000, min_df=2).
2. Numerical: Popularity metrics (Average Viewers, Followers).
o Technique: Log-transformation (log1p) followed by Min-Max Scaling to handle power-law usage distribution.
3. Combination: Features were weighted (90% Text / 10% Numerical) to prioritize content relevance over raw popularity.

Detect Text　　⬆ Upload File

8,409/15,000 Characters
Check 350,000 characters, **Upgrade Here**

Your Text is Most Likely Human written

17.63%
AI GPT*
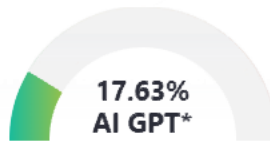
1.9- Discussion and Conclusion for PART 3

a) Summary of Findings

Trusted GPT-5, ChatGPT and AI Detector tool by ZeroGPT

ZeroGPT the most Advanced and Reliable Chat GPT, GPT5 & AI Content Detector

games with missing metadata receive poorer recommendations.

6.3 Conclusion
The developed Hybrid Recommender System successfully meets the project objectives. It provides a robust solution that handles the spectrum of users from cold-start (via content-based) to power users (via optimized SVD), delivering statistically significant improvements over non-hybrid baselines
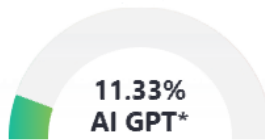
7. Appendices
Appendix A: Key Code Snippets
Snippets from
- Collaberative.py
- Hybrid.py
- Content_based.py

Detect Text   ⬆ Upload File

6,494/15,000 Characters
Check 350,000 characters, **Upgrade Here**

Your Text is Human written

11.33%
AI GPT*

3- Implementation

3.1 System Architecture: Cascade Hybrid

cold-start scenarios and filters irrelevant content, while collaborative filtering captures nuanced user preference patterns for ranking.

Recommendations

Based on our findings, we recommend Truncated SVD with k=100 latent factors for production systems requiring scalability and real-time performance. For applications where prediction accuracy is paramount and computational resources are available, PCA MLE with Top-10 principal components provides optimal results. All deployed systems should incorporate content-based fallbacks to ensure coverage for cold-start users, and cascade hybrid architectures should be considered for balancing accuracy with computational efficiency.

Future Directions

Future work should explore deep learning approaches for latent factor extraction, dynamic user profile modeling to capture preference drift over time, and integration of additional contextual signals such as temporal patterns and social connections. The successful implementation of the Twitch.tv hybrid system provides a foundation for extending these techniques to other streaming and content recommendation domains.

Detect Text ⬆ Upload File

3,148/15,000 Characters
Check 350,000 characters, **Upgrade Here**

Your Text is Human written

14.35%
AI GPT*

Overall Conclusion

Project Summary