

# ДОДАТОК А

## Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



Дата звіту 6/9/2025  
Дата редагування ---



Звіт не був оцінений

### Звіт подібності

#### метадані

Назва організації  
**Kharkiv National University of Radio Electronics**  
Заголовок  
**2025\_М\_ПІ\_ІПЗзм-23-1\_Андрющенко\_Д\_О\_скорочений**  
Автор Науковий керівник / Експерт  
**Андрющенко Дарія Олегівна Олена Олійник**  
підрозділ  
**каф. ПІ**

#### Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



25

Довжина фрази для коефіцієнта подібності 2



6745

Кількість слів

49869

Кількість символів

#### Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про МОЖЛИВІ маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв		0
Інтервали		0
Мікропробіли		0
Білі знаки		0
Парафрази (SmartMarks)		36

#### Подібності за списком джерел

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Колір тексту означає в якому джерелі він був знайдений. Ці джерела і значення Коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

10 найдовших фраз		Копіювати текст
ПОРЯДКОВИЙ НОМЕР	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	60 0.89 %
2	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	38 0.56 %
3	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	37 0.55 %
4	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	37 0.55 %
5	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	35 0.52 %

6	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	34 0.50 %
7	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	33 0.49 %
8	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	21 0.31 %
9	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	21 0.31 %
10	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	20 0.30 %
<b>з бази даних RefBooks (0.00 %)</b>		
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
<b>з домашньої бази даних (0.00 %)</b>		
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
<b>з програми обміну базами даних (0.28 %)</b>		
ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	Абу Ібрагім Рафі Рауфович_Розробка платформи для публікації новин та статей_н.к.Сушинський О.Є_122_Бак_денна 5/8/2025 European University (European University)	13 (1) 0.19 %
2	Оцінка впливу обмежень COVID-19 на поширення інфекції та соціоекономічні показники 3/16/2025 National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute (National Technical University of Ukraine Igor Sikorskyi Kyiv Politech Institute)	6 (1) 0.09 %
<b>з Інтернету (8.21 %)</b>		
ПОРЯДКОВИЙ НОМЕР	ДЖЕРЕЛО URL	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content">https://openarchive.nure.ua/server/api/core/bitstreams/a7b8292d-1583-42eb-a7e5-c7acf3dd2318/content</a>	543 (30) 8.05 %
2	<a href="https://openarchive.nure.ua/server/api/core/bitstreams/72019119-2d0b-45ea-817a-5f77104ff748/content">https://openarchive.nure.ua/server/api/core/bitstreams/72019119-2d0b-45ea-817a-5f77104ff748/content</a>	11 (1) 0.16 %

### Список прийнятих фрагментів (немає прийнятих фрагментів)

ПОРЯДКОВИЙ НОМЕР	ЗМІСТ	КІЛЬКІСТЬ ОДНАКОВИХ СЛІВ (ФРАГМЕНТІВ)
------------------	-------	---------------------------------------

ДОДАТОК Б  
Слайди презентації

Дослідження методів  
рекомендації систем  
на основі  
колаборативної  
фільтрації

ст. гр. ПЗзм-23-1 Андрущенко Д.О.  
Науковий керівник: к.т.н., доцент, проф. каф. ПІ Шубін І.Ю.



23 червня 2025

Рисунок Б.1 – Титульний слайд

## Дослідження

### Актуальність:

- Зростання обсягів інформації в інтернеті потребує більш персоналізованих рекомендацій для користувачів
- Рекомендаційні системи використовуються в таких популярних сьогодні галузях як e-commerce, стримінг, соціальні мережі
- Колаборативна фільтрація – один з найефективніших підходів

### Напрямок дослідження:

- Дослідження методів колаборативної фільтрації для вибору найточніших рішень

### Об'єкт дослідження:

- Методи колаборативної фільтрації



Рисунок Б.2 – Титульний слайд

# Огляд літератури (аналогів)

## Перелік основних джерел та теорій у галузі

- Ricci F., Rokach L., Shapira B. – “Recommender Systems Handbook”.
- Koren Y., Bell R. – методи матричної факторизації.
- He X. – Neural Collaborative Filtering (NCF).
- Adomavicius G. – класифікація методів рекомендацій.
- Sarwar B. – User-based та Item-based KNN.
- Праці українських дослідників: Лобур М. В., Шварц М. Є.

## Зазначення прогалин у наявних дослідженнях

- Недостатньо прикладних порівнянь методів на одному датасеті.
- Бракує аналізу точності з різними метриками подібності.
- Обмежене використання новітніх моделей (NCF) у локальних дослідженнях.
- Не всі підходи враховують адаптацію до масштабів системи.

---

Рисунок Б.3 – Огляд літератури

# Постановка задачі

## Чітке формулювання проблеми:

- Необхідність вибору найточнішого методу колаборативної фільтрації.
- Вплив метрик подібності та кількості сусідів на точність прогнозів.
- Відсутність єдиного підходу для різних типів систем.

## Опис очікуваних результатів:

- Порівняння точності 4 методів: user-KNN, item-KNN, MF, NCF.
- Визначення найкращої метрики подібності для KNN-методів.
- Формування рекомендацій щодо вибору методу для різних сценаріїв застосування.

---

Рисунок Б.4 – Постановка задачі

# Методологія

Практичне дослідження матиме 2 етапи:

1. Порівняння метрик подібності для кожної моделі окремо – для user-KNN та item-KNN.
2. Порівняння найефективніших моделей з першого етапу між собою, а також з реалізаціями PMF (імовірнісна матрична факторизація) та NCF (нейронна колаборативна фільтрація).



5

Рисунок Б.5 – Методологія

## Обрання метрик подібності

Косинусна подібність

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|^2 * \|\vec{b}\|^2}$$

де  $\vec{a} \cdot \vec{b}$  позначає скалярний добуток векторів,  
 $\|\vec{a}\|$  і  $\|\vec{b}\|$  є нормами (довжинами) цих векторів

Кореляція Пірсона

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

де  $X_i$  та  $Y_i$  є значеннями двох змінних,  
 $\bar{X}$  та  $\bar{Y}$  є середніми значеннями змінних  $X$  та  $Y$ .

Індекс Жаккара

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

де  $|A \cap B|$  - це кількість елементів у перетині множин  $A$  та  $B$ ,  $|A \cup B|$  - це кількість елементів в об'єднанні множин  $A$  та  $B$ .

Кореляція Спірмана

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

де  $d_i$  — різниця рангів кожного спостереження в двох наборах даних,  
 $n$  — кількість спостережень.



6

Рисунок Б.6 – Обрання метрик подібності

# Метрики оцінювання дослідження

- MAE - метрика, яка повідомляє нам середню абсолютну різницю між прогнозованими значеннями та фактичними значеннями у наборі даних.

$$MAE = \frac{\sum_{e=1}^N |P_{x,e} - R_{x,e}|}{N}$$

N – загальна кількість елементів у тестовому наборі;

$P_{x,e}$  – прогнозована оцінка користувача x до елемента e;

$R_{x,e}$  – фактична оцінка користувача x до елемента e.

- RMSE – квадратний корінь середньоквадратичної помилки.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{t=n} (y' - y)^2}$$

n – кількість оцінюваних елементів у датасеті,

$y'$  – фактичне значення i-го елемента,

$y$  – прогнозоване значення для i-го елемента,



$(y' - y)^2$  - квадрат відстані між фактичним та прогнозованим значеннями для i-го елемента.

7

Рисунок Б.7 – Метрики оцінювання дослідження

## Опис програмного забезпечення, що було використано у дослідженні

- мова програмування – Java;
- середовище розробки - IntelliJ Idea;
- бібліотека CF4J на Java.



8

Рисунок Б.8 – Опис програмного забезпечення, що було використано у дослідженні

# Зміст проведеного експерименту

- **Обрані моделі:** user-KNN, item-KNN, collaborative filtering, PMF, NCF.
- **Вхідні дані:** датасет MovieLens 100k.
- **Метрики подібності:** косинусна подібність, кореляція Пірсона, індекс Жаккара, кореляція Спірмана.
- **Критерії оцінювання:** показники похибок MAE та RMSE.
- **Діапазон вимірювання:** від 10 до 50 сусідів (для NCF та PMF - прихованих факторів) з кроком 10.

Рисунок Б.9 – Зміст проведеного експерименту

## Результати дослідження метрик подібності User-KNN модель

Найефективніша метрика подібності для user-KNN моделі виявилась кореляція Спірмана.

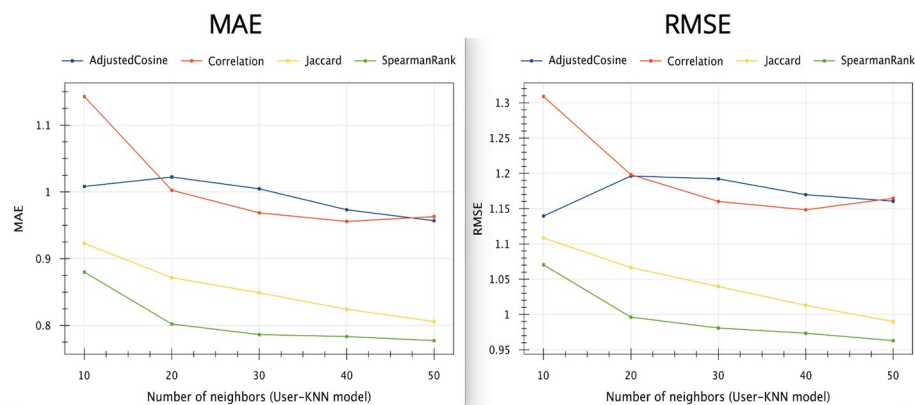


Рисунок Б.10 – Результати дослідження метрик подібності

# Результати дослідження метрик подібності

## Item-KNN модель

Найефективніша метрика подібності для item-KNN моделі – індекс Жаккара.

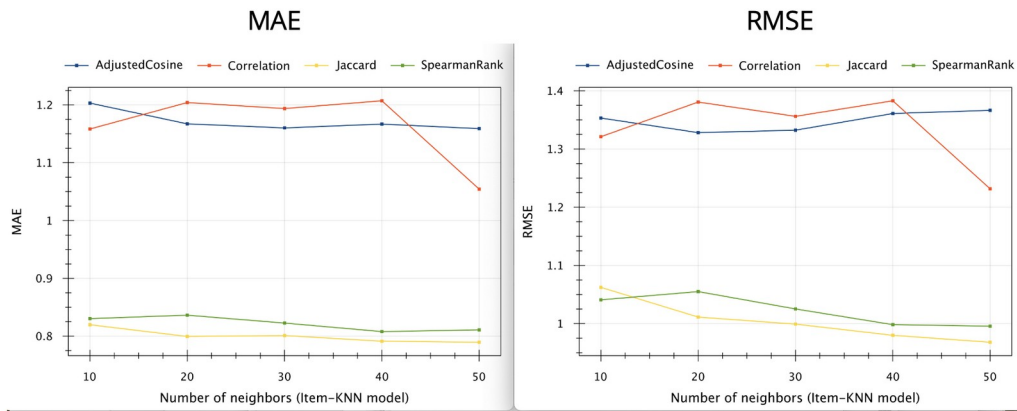


Рисунок Б.11 – Результати дослідження метрик подібності

## Результати експерименту

- для моделі NCF значення похибок є набагато більшими за інші моделі
- для PMF моделі на протязі усього діапазону значення похибок потроху, але стабільно іде вгору

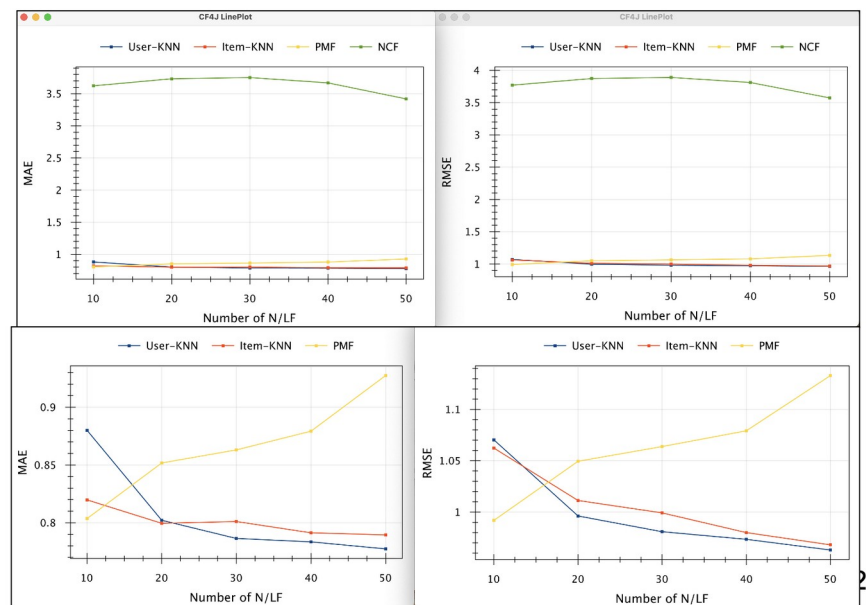


Рисунок Б.12 – Результати експерименту

## Абсолютні значення похибок

Number of N/LF	User-KNN	Item-KNN	PMF	NCF
10	0.8800	0.8198	0.8037	3.1453
20	0.8022	0.7996	0.8518	3.6075
30	0.7864	0.8011	0.8631	3.4275
40	0.7835	0.7913	0.8793	3.4654
50	0.7774	0.7895	0.9274	3.5744

Number of N/LF	User-KNN	Item-KNN	PMF	NCF
10	1.0704	1.0623	0.9919	3.3168
20	0.9962	1.0113	1.0495	3.7533
30	0.9808	0.9992	1.0639	3.5802
40	0.9734	0.9800	1.0792	3.6153
50	0.9630	0.9680	1.1330	3.7199

User-KNN та item-KNN мають дуже схожі значення і в кінці діапазону відрізняються менше ніж на 0,01. Видно, що item-KNN має трохи більшу похибки, ніж user-KNN, але ця різниця незначна, тому при виборі серед цих моделей краще звертати увагу в першу чергу на предметну галузь рекомендаційної системи, для розуміння для якої з цих моделей буде у системи більше інформації.

Рисунок Б.13 – Абсолютні значення похибок

## Аналіз отриманих результатів

- NCF моделі можуть бути неефективними на невеликому наборі даних
- PMF моделі потроху стають менш точними при масштабуванні
- В моделях з метриками подібності найважливіше підібрати найефективнішу
- При виборі між user-KNN та item-KNN моделями треба в першу чергу звертати увагу на домен рекомендаційної системи і вибирати ту систему, для якої домен зможе дати найбільше інформації

Рисунок Б.14 – Аналіз отриманих результатів

# Публікація результатів



## Актуальні питання та перспективи проведення наукових досліджень

Андріющенко Дарія Олегівна, здобувач вищої освіти  
факультету комп'ютерних наук  
Національний університет радіоелектроніки, Харків, Україна

Науковий керівник: Шубін Ігор Юрійович, проф. каф. програмної інженерії,  
канд. техн. наук, доцент  
Національний університет радіоелектроніки, Харків, Україна

## КРИТЕРІЇ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ АЛГОРИТМІВ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ

Оцінка ефективності алгоритмів колаборативної фільтрації є важливою складовою для розробки високоякісних рекомендаційних систем. Для цього використовуються різні критерії, що дозволяють детально оцінити точність, охоплення, швидкість та здатність системи адаптуватися до змін. Серед таких критеріїв можна виділити п'ять основних: RMSE, Coverage, Time, Novelty та Stability.

RMSE (Root Mean Square Error) – це один із основних критеріїв для вимірювання точності рекомендаційної системи. Він оцінює точність передбачених оцінок шляхом вимірювання середнього квадратичного відхилення між фактичними та передбаченими оцінками. RMSE є важливим, оскільки він дає зрозуміти, наскільки точно алгоритм передбачає вподобання користувачів. Чим менше значення RMSE, тим точнішими є рекомендації, що генерує система [1]. Це дозволяє оцінити, наскільки система здатна надавати коректні рекомендації на основі наявних даних.

15

Рисунок Б.15 – Публікація результатів

## Підсумки

### Дослідження демонструє, що:

- більш прості і традиційні методи можуть бути більш ефективними у певних сценаріях
- важливість правильного вибору та налаштування параметрів рекомендаційних систем
- важливість врахування домену рекомендаційної системи при виборі моделі прогнозування

### Подальші дослідження:

- дослідження гібридних моделей
- дослідження інших реалізацій матричної факторизації та нейронних мереж
- збільшення діапазону дослідження



16

Рисунок Б.16 – Підсумки

## ДОДАТОК В

### Апробація результатів роботи

**Актуальні питання та перспективи проведення наукових досліджень**

**Андрющенко Дарія Олегівна**, здобувач вищої освіти  
факультету комп'ютерних наук  
*Національний університет радіоелектроніки, Харків, Україна*

**Науковий керівник: Шубін Ігор Юрійович**, проф. каф. програмної інженерії,  
канд. техн. наук, доцент  
*Національний університет радіоелектроніки, Харків, Україна*

### **КРИТЕРІЇ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ АЛГОРИТМІВ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ**

Оцінка ефективності алгоритмів колаборативної фільтрації є важливою складовою для розробки високоякісних рекомендаційних систем. Для цього використовуються різні критерії, що дозволяють детально оцінити точність, охоплення, швидкість та здатність системи адаптуватися до змін. Серед таких критеріїв можна виділити п'ять основних: RMSE, Coverage, Time, Novelty та Stability.

RMSE (Root Mean Square Error) – це один із основних критеріїв для вимірювання точності рекомендаційної системи. Він оцінює точність передбачених оцінок шляхом вимірювання середнього квадратичного відхилення між фактичними та передбаченими оцінками. RMSE є важливим, оскільки він дає зрозуміти, наскільки точно алгоритм передбачає вподобання користувачів. Чим менше значення RMSE, тим точнішими є рекомендації, що генерує система [1]. Це дозволяє оцінити, наскільки система здатна надавати коректні рекомендації на основі наявних даних.

Coverage — цей критерій визначає відсоток об'єктів або предметів, для яких алгоритм здатен надати рекомендацію. Coverage оцінює здатність системи охоплювати весь асортимент доступних предметів і давати рекомендації на їх основі [1]. Високе значення coverage означає, що система здатна рекомендувати більшу кількість предметів користувачам. Однак важливо, щоб цей критерій поєднувався з іншими метриками, як-то точність, оскільки просто велика кількість рекомендацій не гарантує їх коректність або релевантність для користувача.

Time — це критерій, який вимірює середній час генерації рекомендацій. Оскільки багато рекомендаційних систем працюють в реальному часі, швидкість їхнього виконання є важливою характеристикою. Час, необхідний для надання рекомендацій, впливає на досвід користувача та ефективність системи. Чим менше цей час, тим краще для системи, оскільки це забезпечує швидку реакцію на запити користувачів. Однак варто зазначити, що інколи оптимізація часу може впливати на точність, тому баланс між точністю та швидкістю є важливим для ефективної роботи алгоритму [2].

Novelty — цей критерій оцінює здатність алгоритму рекомендувати нові, менш популярні або незвичні елементи, які користувач, ймовірно, не знайшов би самостійно. Novelty є важливим для підтримки інтересу користувачів, оскільки дає їм можливість відкривати нові предмети, які можуть їм сподобатися, але вони б могли бути не знайдені через високу популярність інших елементів. Алгоритми, які забезпечують високу новизну, здатні рекомендувати елементи, що є більш

індивідуальними та персоналізованими для користувачів, що, в свою чергу, покращує задоволення користувачів і сприяє розвитку інтересу до нових продуктів чи послуг.

Stability — цей критерій оцінює стійкість результатів рекомендаційної системи при зміні даних, таких як додавання нового користувача або нового предмета. Алгоритм повинен бути здатний адаптуватися до змін у даних без суттєвих коливань у якості рекомендацій. Наприклад, система не повинна сильно змінювати рекомендації для існуючих користувачів лише через додавання нового елемента чи користувача в базу даних. Висока стабільність означає, що алгоритм забезпечує сталі результати навіть у випадку внесення змін або розширення системи, що є важливим для надійності роботи рекомендаційної системи.

Таким чином, ці п'ять критеріїв — RMSE, Coverage, Time, Novelty і Stability — є основними показниками для оцінки ефективності алгоритмів колаборативної фільтрації. Вони дозволяють всебічно оцінити роботу рекомендаційних систем, зважаючи на точність, охоплення, швидкість, здатність до новизни та стабільність, що сприяє створенню більш точних, швидких та персоналізованих рекомендацій.

**Список використаних джерел:**

1. Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5-53 с.
2. Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to Recommender Systems Handbook*. Springer.

Рисунок В.2 – Друга сторінка статті

# ДОДАТОК Г

## Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент  
(посада)

програмної інженерії  
(кафедра)

ІПЗм-23-1  
(група)

Андрющенко Дарія Олегівна

(прізвище, ім'я, по батькові)

### Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	<b>7.1 Загальні положення</b>	
	<b>7.3 Нумерація сторінок звіту</b>	
	<b>7.4 Нумерація розділів, підрозділів, пунктів, підпунктів</b>	
	<b>7.5 Рисунки</b>	
	<b>7.6 Таблиці</b>	
	<b>7.7 Переліки</b>	
	<b>7.8 Примітки</b>	
	<b>7.9 Виноски</b>	
	<b>7.10 Формули та рівняння</b>	
	<b>7.11 Посилання</b>	
	<b>7.13 Список авторів</b>	
	<b>7.14 Скорочення та умовні позначки</b>	
	<b>7.15 Додатки</b>	

зауважень немає

Експерт

\_\_\_\_\_  
(підпис)

Олена ОЛІЙНИК  
(прізвище, ініціали)

11.06.2025