

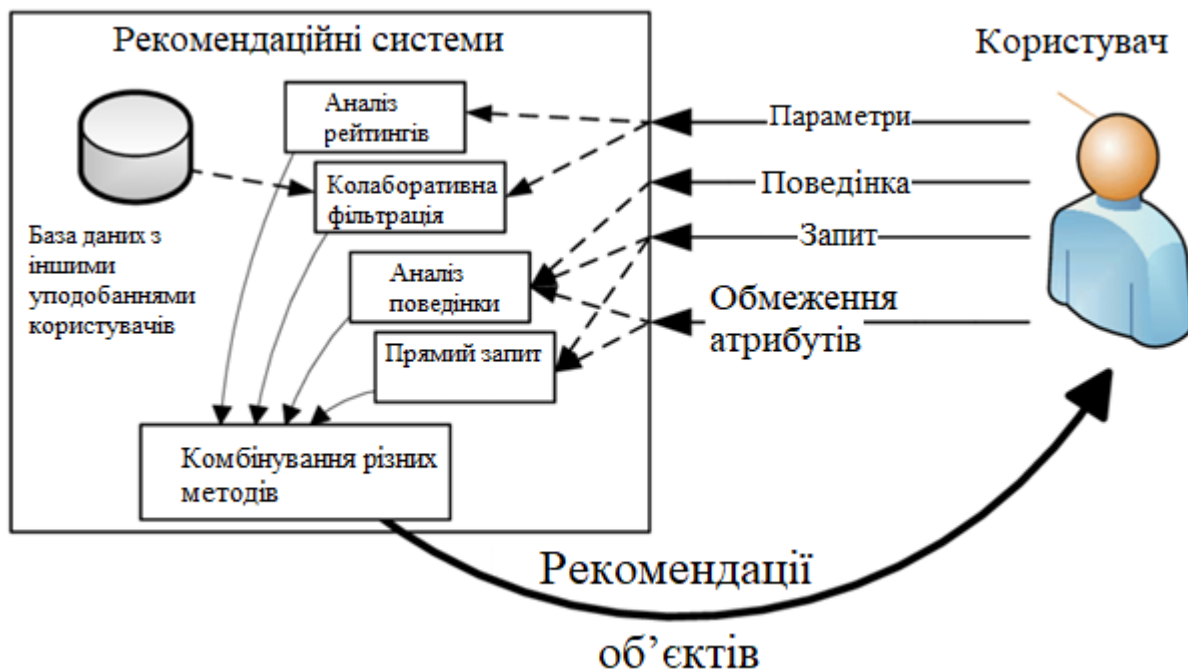
## **ДОДАТОК А**

Графічний матеріал

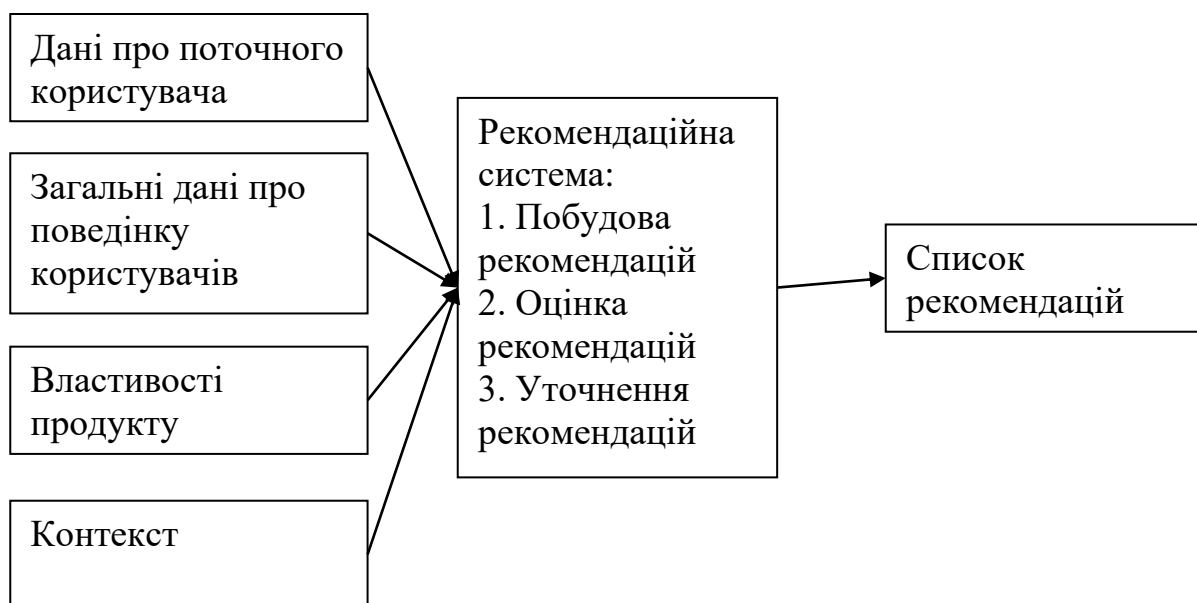
## ЦІЛІ ТА ЗАДАЧІ ДОСЛІДЖЕННЯ

Актуальність роботи	На даний момент існує множина методів для формування рекомендацій, але всі вони мають свої переваги і недоліки. Саме тому дослідження в даній області актуальні. Проблема особливо актуальна для нових, щойно створених систем.
Об'єкт дослідження	Процес побудови рекомендацій.
Предмет дослідження	Методи оцінки рекомендаційних систем.
Мета роботи	Підвищення ефективності побудови формування рекомендацій на основі ітеративної оцінки результатів роботи рекомендаційних систем.
Задачі дослідження	<ol style="list-style-type: none"> <li>1. Аналіз методів оцінки рекомендаційних систем і постановка задачі;</li> <li>2. Дослідження методів оцінювання рекомендацій з використанням явного та неявного зворотнього зв'язку;</li> <li>3. Структурно-функціональне моделювання процесу оцінювання рекомендаційних систем;</li> <li>4. Інтеграція побудови рекомендацій та оцінювання їх результатів;</li> <li>5. Експериментальна перевірка удосконаленого методу.</li> </ol>
Научна новизна	Удосконалений метод колаборативної фільтрації з неявним зворотнім зв'язком, сортування засноване на попарному порівнянні переваг кожного користувача по відношенню до об'єктів у матриці вихідних даних. У результаті сформувався порядок елементів кожного користувача, що дозволило повисити точність методів оцінки рекомендаційних систем.
Практичні результати	Підвищення релевантності рекомендацій для користувача в рекомендаційній системі.

## СТРУКТУРА РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ



## ПРОЦЕС РОБОТИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ



## **ДОСЛІДЖЕННЯ МЕТОДІВ ОЦІНЮВАННЯ РЕКОМЕНДАЦІЙ З ВИКОРИСТАННЯМ ЯВНОГО ТА НЕЯВНОГО ЗВОРОТНЬОГО ЗВ'ЯЗКУ**

Алгоритм формування рекомендації використовує набір даних про вибір користувача для обчислення моделі користувацьких переваг. Потім ця модель може бути використана для обчислення рекомендацій для користувачів системи.

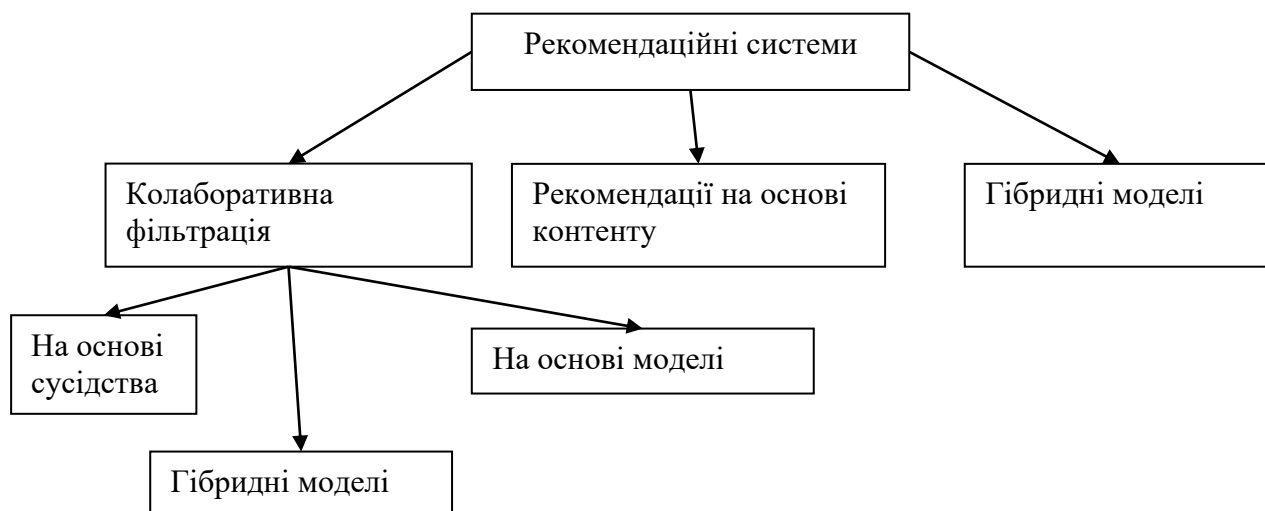
Методи оцінки рекомендаційних систем:

- експериментальні ("Онлайн-експерименти", тестування, рандомізоване контрольоване дослідження, принаймні, одна змінна маніпулюється, одиниці виміру призначаються випадково.

- не експериментальні ("автономні експерименти", історичні дані)

- імітаційний експеримент

## ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ПІДБОРУ РЕКОМЕНДАЦІЙ

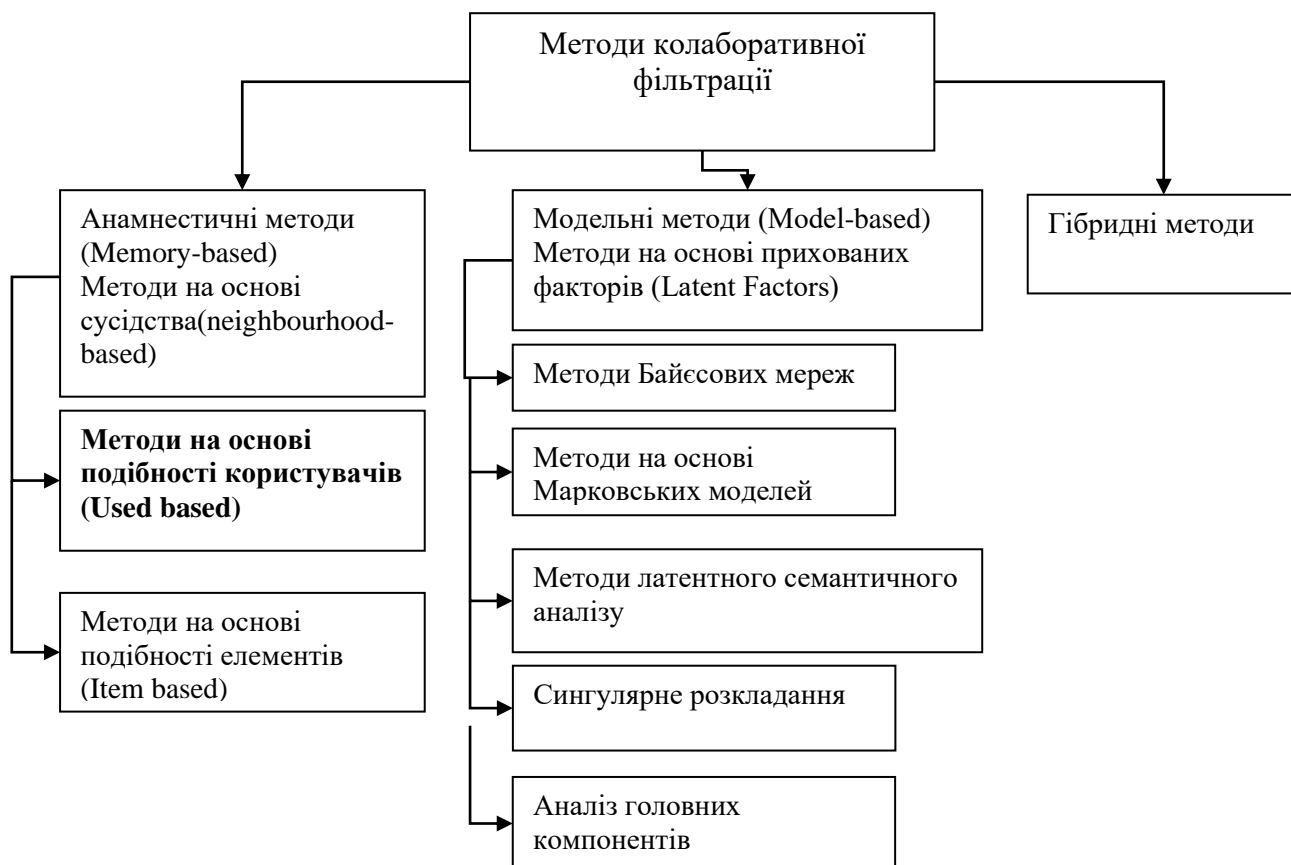


	Контекстна фільтрація	<b>Колаборативна фільтрація</b>	Фільтрація на знаннях	Гібридна фільтрація
Проблема холодного старту	Відсутня	Присутня	Відсутня	Відсутня
Складність реалізації	Низька	Низька	Середня	Висока
Специфіка роботи	Інтернет-магазини, портали кіно, музики та ін.	Інтернет-магазини, портали кіно, музики та ін.	Інтернет-магазини	Будь-яка область
Точність рекомендацій	Середня	Середня	Висока	Висока
Потенційна швидкість роботи	Висока	Середня	Середня	Висока
Залежність від інших користувачів	Відсутня	Присутня	Відсутня	Залежить від реалізації

## ПРОБЛЕМА ДОСЛІДЖЕННЯ

Для того, щоб система мала високоякісні рекомендації, необхідно постійно оновлювати інформацію про інтереси користувача. Але існує й інша проблема – користувач завжди хоче мінімізувати час, що витрачається на взаємодію з системою, і неохоче ділиться інформацією для зворотного зв'язку з системою. Це типова проблема для всіх рекомендаційних систем, і оскільки рекомендації засновані виключно на інформації профілю користувача, тобто чим менше інформації користувач надає, тим менше відповідний набір рекомендацій користувач отримає. Сучасні методи підготовки рекомендацій враховують властивості предметів а також користувачів, які вибирають ці предмети. Однак оцінювання ефективності цих методів виконується лише при їх реалізації. На практиці існує потреба у регулярній адаптації цих методів на основі оцінки їх ефективності. Це визначає актуальність теми дослідження.

## КЛАСИФІКАЦІЯ МЕТОДІВ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ



В якості вхідних даних для КФ використовується надана користувачами інформація про переваги одних предметів над іншими. Перевага (оцінка) часто представляється у вигляді триплета (користувач, предмет, оцінка). Ці оцінки можуть приймати різні форми, в залежності від конкретної системи. Множина всіх триплетів оцінок формує розріджену матрицю, звану матрицею оцінок. Пари (користувач, предмет), в яких користувачі не віддали перевагу предмету, є невідомими значеннями цієї матриці:

	Елемент 1	Елемент 1	Елемент 1
Користувач 1	3	?	2
Користувач 2	?	4	3
Користувач 3	5	4	?

## ОПИС УДОСКОНАЛЕННЯ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ НА ОСНОВІ ПРОПУЩЕНИХ ДАНИХ З НЕЯВНИМ ЗВОРОТНИМ ЗВ'ЯЗКОМ

В якості вихідних даних при побудові рекомендацій використовуються:

- список користувачів  $U = \{u_i\}$ ;
- список об'єктів, що цікавлять користувачів  $E = \{e_j\}$ , що побудовані за результатом колаборативної фільтрації;
- матриця рейтингів/покупок  $R \subseteq U \times E$ ;

Користувачі взаємодіють з об'єктами (товарами), формуючи матрицю рейтингів в рекомендаційній системі:  $R = \{r_{ij}\}$ , де  $R_{ij}$  – рейтинг  $j$  – предмета  $e_j$  у користувача  $u_i$ .

Етап 1: для кожного елемента  $j$  обчислюється міра близькості до елемента  $i$  на основі коефіцієнта Пірсона:

$$s_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \quad (1)$$

де  $U \in U_j \cup U_i$  – множина користувачів, які оцінили елементи  $i$  та  $j$ .

Етап 2: вибираємо множина елементів  $S$ , найбільш близьких до об'єкта  $i$ . Відомо, що достатні результати виходять при  $k=30$  елементів множини  $S$ . Але ці дані залежні від розглянутої задачі і розрідженості матриці.

Етап 3: прогноз рейтингу (оцінки) об'єкта на основі рейтингів близьких до нього об'єктів:

$$\hat{r}_{u,i} = \frac{\sum_{j \in S} s_{i,j} \cdot r_{u,j}}{\sum_{j \in S} |s_{i,j}|} \quad (2)$$

На основі представлених математичних викладок були проведені дослідження з використання описаного методу. Для оцінки ефективності використовувалася середньоквадратичне відхилення.

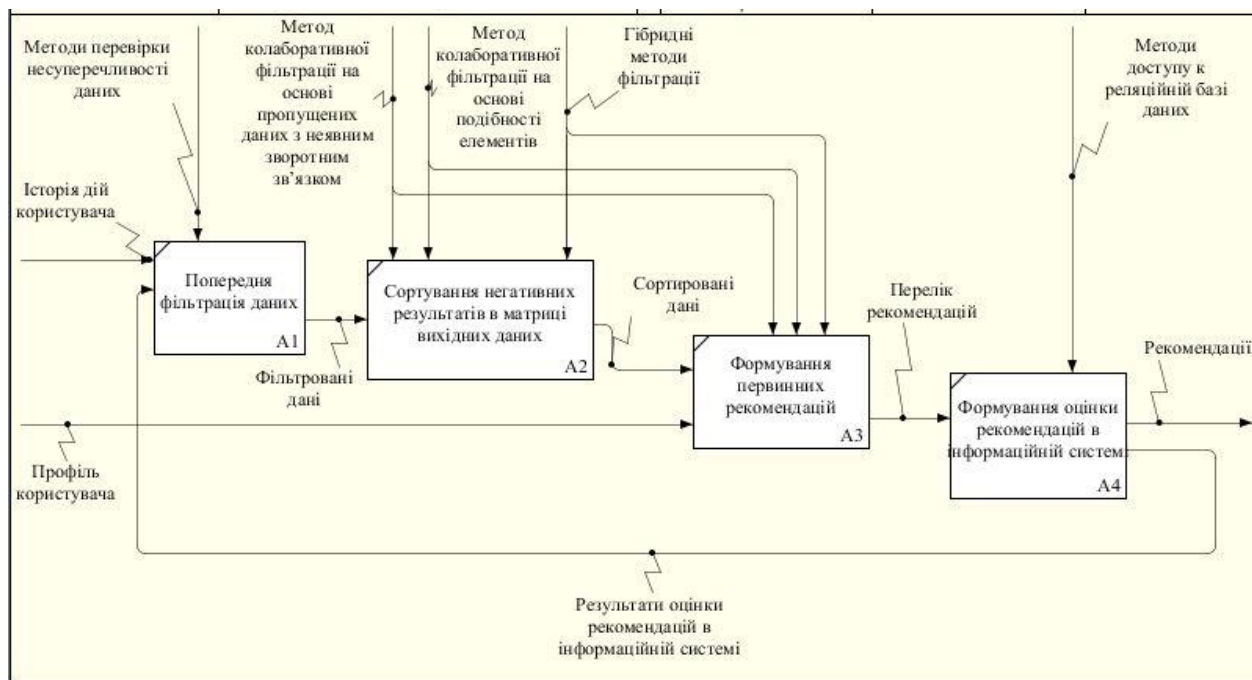
$$RMSE = \sqrt{\frac{1}{n} \sum_{u,i} (\hat{r}_{u,i} - r_{u,i})^2} \quad (3)$$

Де  $r_{u,i}$  – відома оцінка користувача  $u$  для елемента  $i$ ,  $\hat{r}_{u,i}$  – прогнозована оцінка.

	$e_1$	$e_2$	$e_3$	$e_4$
$u_1$	0	9	6	0
	$e_1$	$e_2$	$e_3$	$e_4$
$e_1$	=	9	6	0
$e_2$	+	=	+	+
$e_3$	+	-	=	+
$e_4$	?	-	-	=

На даному кроці виконується попарне порівняння результатів (покупок, переглядів) для кожного користувача. Для користувача формується квадратна матриця, елементи якої вказують, який з об'єктів є більш цікавим для нього. Знак  $+$  в цій матриці означає, що елемент в рядку більш цікавий, ніж елемент в стовпці. Приклад реалізації даного кроку для користувача 1 у наведено на рис.

## СТРУКТУРНО-ФУНКЦІОНАЛЬНЕ МОДЕЛЮВАННЯ ПРОЦЕСУ ОЦІНЮВАННЯ РЕКОМЕНДАЦІЙНИХ СИСТЕМ



## РЕАЛІЗАЦІЯ УДОСКОНАЛЕНОГО МЕТОДУ

Вхідні дані передаються в програму шляхом завдання аргументів для головного методу main серверної частини.

```
def buildingArtistByID(): DataFrame = {  
  rawArtistsData.flatMap { line =>  
    val (id, name) = line.span(_ != '\t')  
    if (name.isEmpty) {  
      None  
    } else {  
      try {  
        Some((id.toInt, name.trim))  
      } catch {  
        case _: NumberFormatException => None  
      }  
    }  
  }  
}
```

Приклад вхідних даних:

ID виконавця	Назва виконавця
1195	Elvis Costello
1004346	Joshua Redman
658	George Duke
1003472	Brian Hughes
3437	Enigma

## ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА УДОСКОНАЛЕНОГО МЕТОДУ

### Результати AUC

Початкова величина AUC	0,862
Величина AUC за результатами корегування даних після першої оцінки ефективності	0,877
Величина AUC за результатами корегування даних після другої оцінки ефективності	0,881

Були отримані схожі результати. Для порівняння було проведено налаштування параметрів алгоритму колаборативної фільтрації. Оцінки вийшли аналогічними. Але у першому варіанті можна автоматизувати процес.

### AUC при різних гіперпараметрах

Величина AUC	Значення гіперпараметрів		
	Lambda	Alpha	Iterations
0.889	5.0	40.0	20
0.883	1.0	1.0	20
0.840	1.0	1.0	10
0.800	1.0	1.0	20
0.793	5.0	1.0	20
0.733	1.0E-4	40.0	10
0.694	1.0E-4	15.0	10
0.624	1.0E-4	40.0	10

## ВИСНОВКИ

Сервіси збирають інформацію про переваги користувачів і намагаються запропонувати їм корисні товари. На даний момент існує множина методів для формування рекомендацій, але всі вони мають свої переваги і недоліки. Саме тому дослідження в даній області були актуальні.

У роботі виконано дослідження методів оцінки рекомендаційних систем:

- проаналізовано існуючі методи оцінки;
- на підставі проведеного аналізу запропоновано удосконалення оцінювання рекомендаційних систем шляхом оцінювання переваг об'єктів для кожного користувача метод колаборативної фільтрації з неявним зворотнім зв'язком. Метод передбачає сортування об'єктів на основі попарного порівняння переваг кожного користувача по відношенню до об'єктів у матриці вихідних даних

Виконана експериментальна перевірка.

В ході дослідження отримані такі результати: визначені існуючі рекомендаційні системи, виконано аналіз процесу роботи рекомендаційних систем, аналіз критеріїв та методів оцінювання рекомендаційних систем, побудовано удосконалену рекомендаційну систему за на основі колаборативної фільтрації, побудовано контекстну діаграму моделі «Дослідження методів оцінки рекомендаційних систем», проведено експериментальну перевірку удосконаленого методу.