

# Дослідження методів виявлення спільних вподобань користувачів

Автор: Метелиця Марія Андріївна

Науковий керівник: Кухарев Сергій Олександрович

# Актуальність

- Зростання кількості Інформації
- Зростання кількості користувачів
- Розвиток соціальних мереж
- Комерціалізація інновацій

## **Мета роботи:**

- аналіз та порівняння методів виявлення спільних вподобань групи користувачів;
- практична реалізація групової рекомендаційної системи на основі досліджених методів

**Об'єкт дослідження:** користувацькі профілі

**Предмет дослідження:** методи виявлення спільних вподобань користувачів

# Постановка задачі

## Користувацькі профілі

Профіль користувача 1		Профіль користувача 2		Профіль користувача N	
Об'єкт	Оцінка	Об'єкт	Оцінка	Об'єкт	Оцінка
A11	8	A21	6	AN1	10
...	...	...	...	...	...
A1L	6	A2M	2	ANK	8

## Групова рекомендація

Об'єкт	Оцінка
A1	9
...	...
AZ	2

# Сучасний стан проблеми

## **Методи виявлення спільних вподобань:**

- Агрегування вихідних користувацьких профілів (Model aggregation)
- Агрегування передбачень для окремих користувачів (Prediction aggregation)

## **Приклади групових рекомендаційних систем:**

MUSICFX, INTRIGUE, POLYLENS, TRAVEL DECISION FORUM

# Виявлення індивідуальних вподобань

## Колаборативна фільтрація

Коефіцієнт кореляції Пірсона

$$w_{u,w} = \frac{\sum_i (r_{u,i} - \bar{r}_u)(r_{w,i} - \bar{r}_w)}{\sqrt{\sum_i r_{u,i}^2} \sqrt{\sum_i r_{w,i}^2}},$$

де  $r_{u,i}$  – рейтинг, який користувач  $u$   
поставив об'єкту  $i$

Рейтинг користувача  $u$

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum (r_{w,i} - \bar{r}_w) w_{u,w}}{\sum_w |w_{u,w}|}$$

# Виявлення спільних вподобань

Утилітарна адитивна

$$gr_m = \sum_{n=1}^N r_{n,m}$$

Утилітарна мультиплікативна

$$gr_m = \prod_{n=1}^N r_{n,m}$$

Стратегія Середнього

$$gr_m = \frac{1}{N} \sum_{n=1}^N r_{n,m}$$

Стратегія Середнього без страждань

$$gr_m = \begin{cases} \frac{1}{N} \sum_{n=1}^N r_{n,m}, & r_{n,m} \geq T \\ 0, & r_{n,m} < T \end{cases}$$

Стратегія Мінімальних страждань

$$gr_m = \min_{n=1,N} r_{n,m}$$

Стратегія Максимального задоволення

$$gr_m = \max_{n=1,N} r_{n,m}$$

- Стратегія Справедливості
- Стратегія Борда
- Стратегія правило Копленда
- Стратегія Голосування за згодою
- Мажоритарна стратегія

де  $gr_m$  - груповий рейтинг для **m-го** фільму;  $N$  – кількість користувачів у групі;  
 $r_{n,m}$  - рейтинг **n-го** користувача **m-му** фільму;  $T$  – порогове значення.

# Метрики оцінювання точності

## Середнє абсолютне відхилення для групи

- за принципом максимальної похибки

$$MAE_{gr}^{max} = \frac{1}{M_n} \max_{n=1, N} \left\{ \sum_{m=1}^{M_n} |r_{nm} - gr_m| \right\}$$

- за принципом середнього арифметичного

$$MAE_{gr}^{av} = \frac{1}{NM_n} \sum_{n=1}^N \sum_{m=1}^{M_n} |r_{nm} - gr_m|$$

## Середньоквадратична похибка для групи

- за принципом максимальної похибки

$$RMSE_{gr}^{max} = \max_{n=1, N} \left\{ \sqrt{\frac{1}{M_n} \sum_{m=1}^{M_n} (r_{nm} - gr_m)^2} \right\}$$

- за принципом середнього арифметичного

$$RMSE_{gr}^{av} = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{M_n} \sum_{m=1}^{M_n} (r_{nm} - gr_m)^2}$$

де  $M_n$  – кількість фільмів, які оцінив n-й користувач;  
 $r_{nm}$  – вихідний рейтинг m-го фільму для n-го користувача;  
 $gr_m$  – отриманий груповий рейтинг для m-го фільму.

# Метрика оцінювання ранжування

- Групова оцінка ранжування за принципом середнього арифметичного

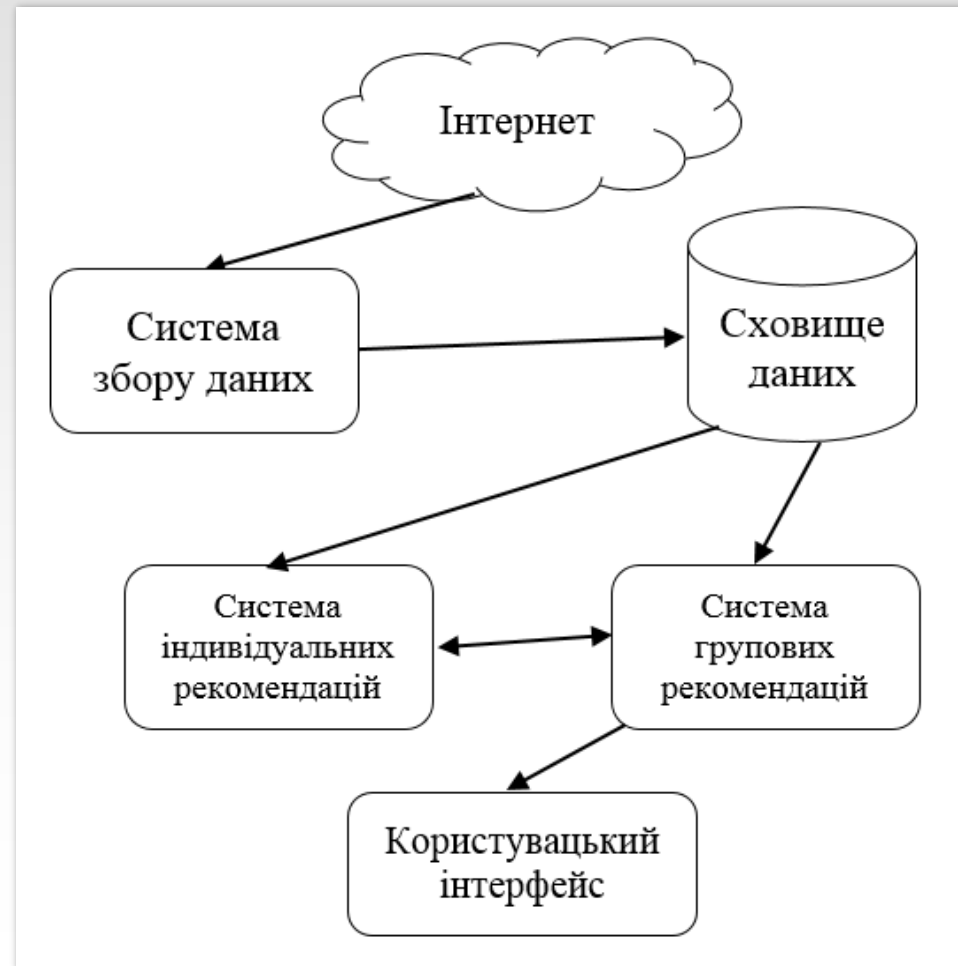
$$RankScore_{gr}^{av} = \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^{M_n} \frac{r_{nm} - \bar{r}_n}{2^{rank_m/M}}$$

- Групова оцінка ранжування за принципом мінімального задоволення

$$RankScore_{gr}^{min} = \min_{n=1,N} \left\{ \sum_{m=1}^{M_n} \frac{r_{nm} - \bar{r}_n}{2^{rank_m/M}} \right\}$$

де  $rank_m$  – місце в загальному рейтингу фільмів для групи;  
 $r_{nm}$  - рейтинг, який  $n$ -й користувач поставив  $m$ -му фільму;  
 $\bar{r}_n$  – середній рейтинг для  $n$ -го користувача  $\bar{r}_n = \frac{1}{M_n} \sum_{m=1}^{M_n} r_{nm}$ ;  
 $M$  – загальна кількість фільмів у матриці вихідних даних.

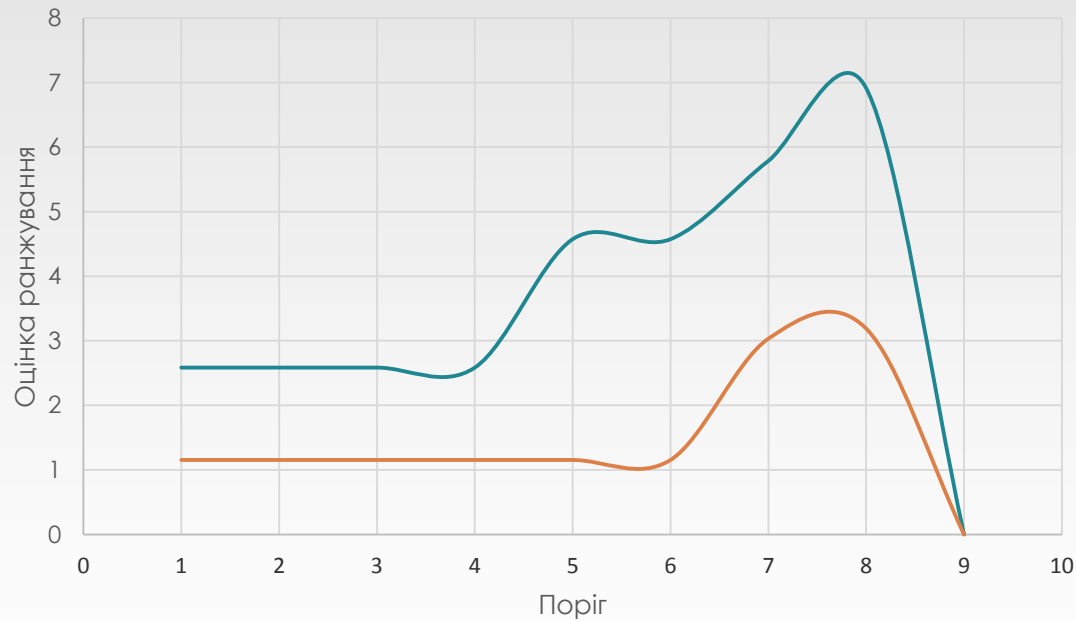
# Схема групової рекомендаційної системи



# Аналіз практичних результатів

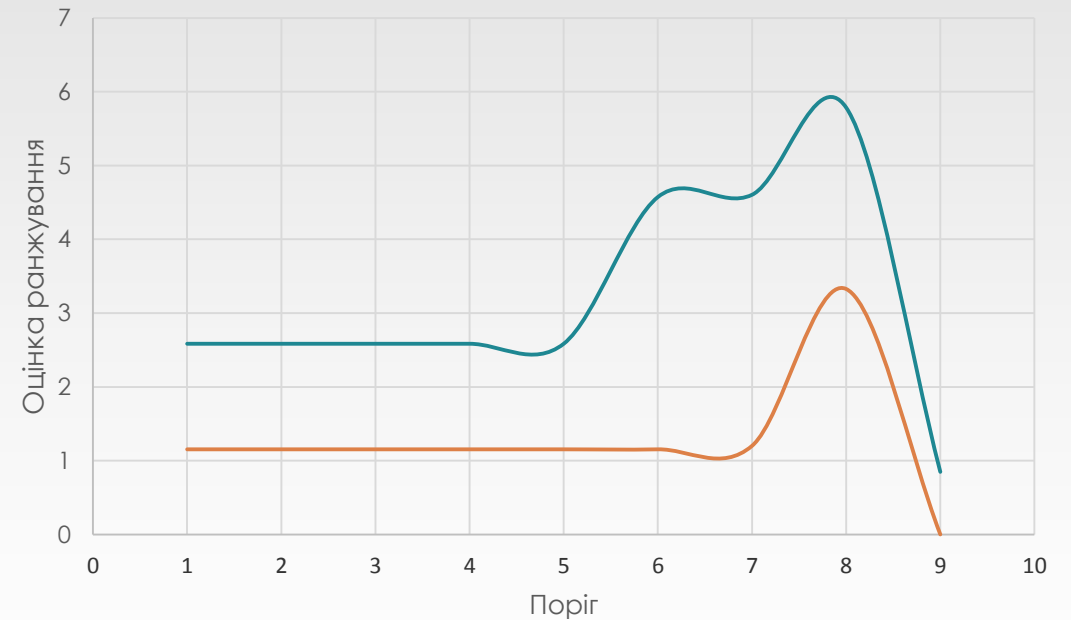
# Дослідження впливу параметра T

## Стратегія Середнього Без Страждань



— Оцінка ранжування за принципом середнього арифметичного  
— Оцінка ранжування за принципом мінімального задоволення

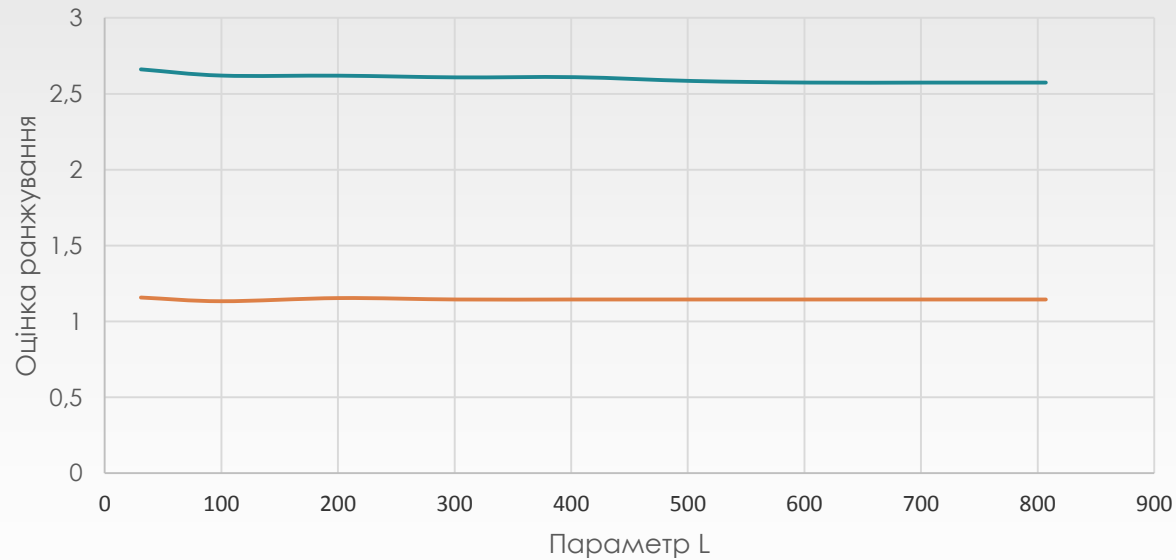
## Стратегія Голосування За Згодою



— Оцінка ранжування за принципом середнього арифметичного  
— Оцінка ранжування за принципом мінімального задоволення

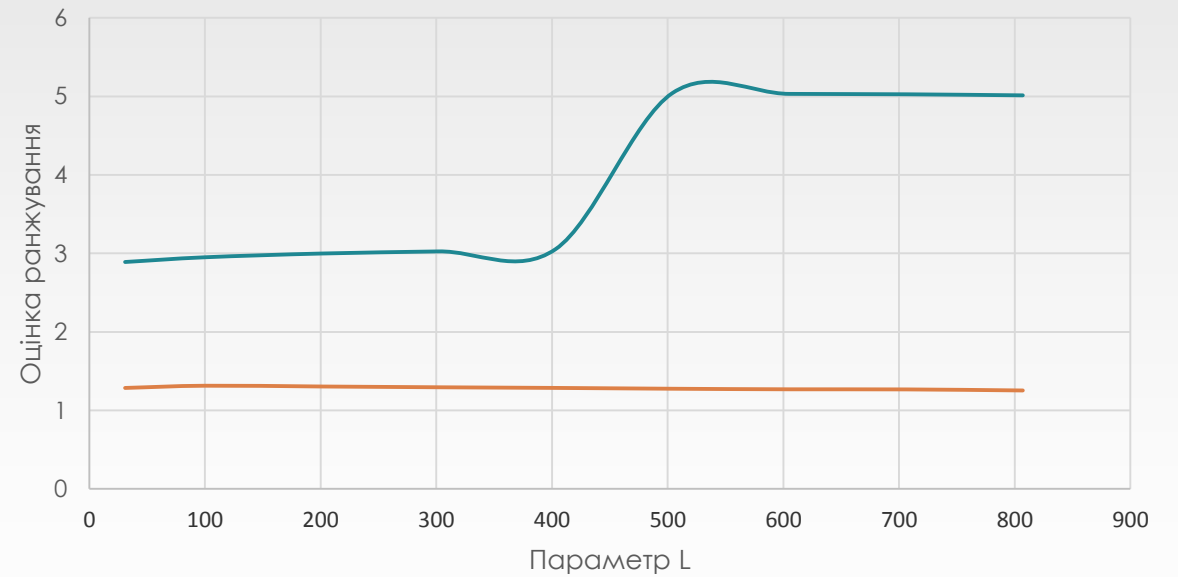
# Дослідження впливу параметра L

## Стратегія Більшості Голосів



- Оцінка ранжування за принципом середнього арифметичного
- Оцінка ранжування за принципом мінімального задоволення

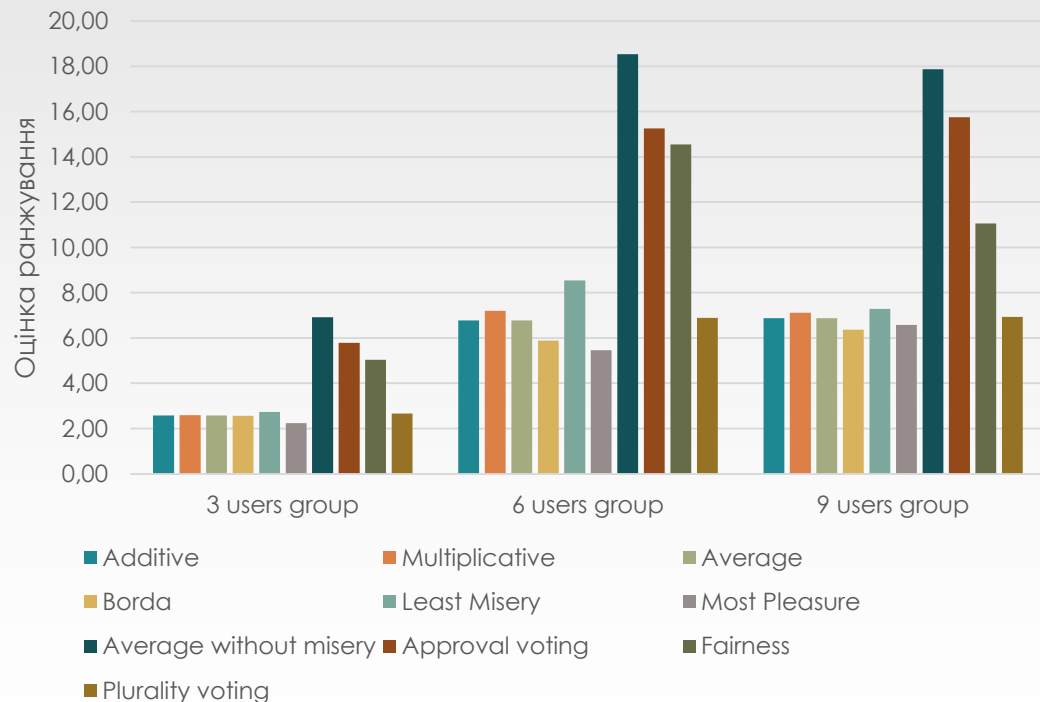
## Стратегія Справедливості



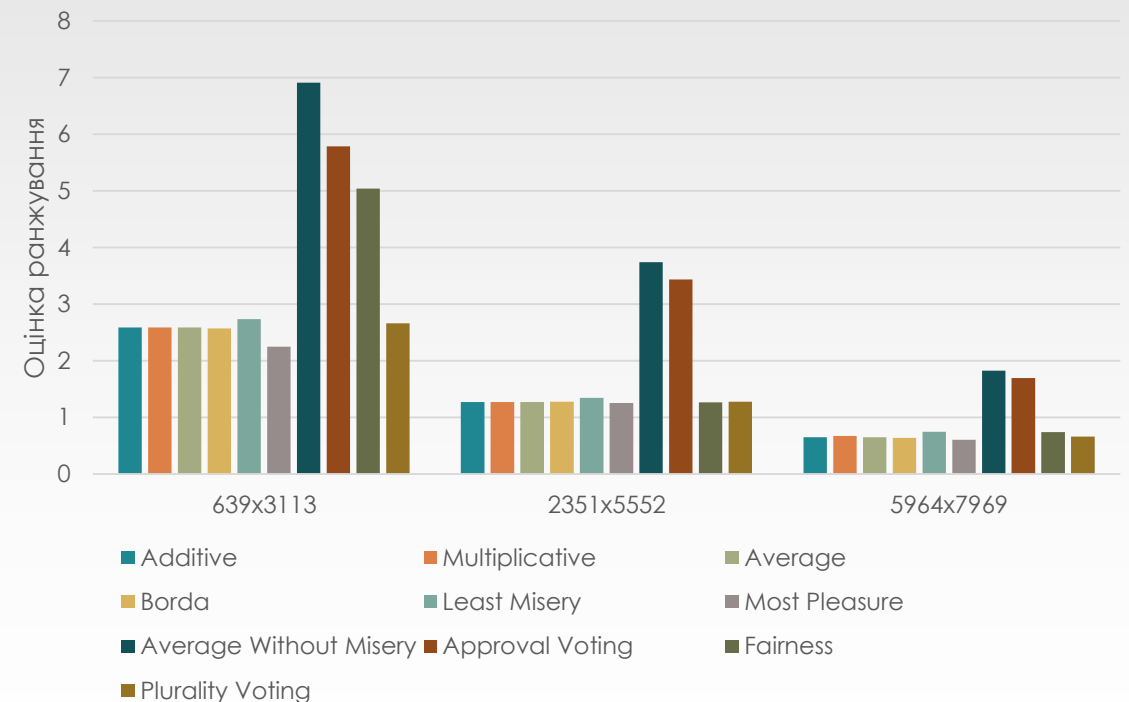
- Оцінка ранжування за принципом середнього арифметичного
- Оцінка ранжування за принципом мінімального задоволення

# Порівняння стратегій суспільного вибору за оцінкою ранжування для груп та вибірок різного розміру

Стратегії суспільного вибору

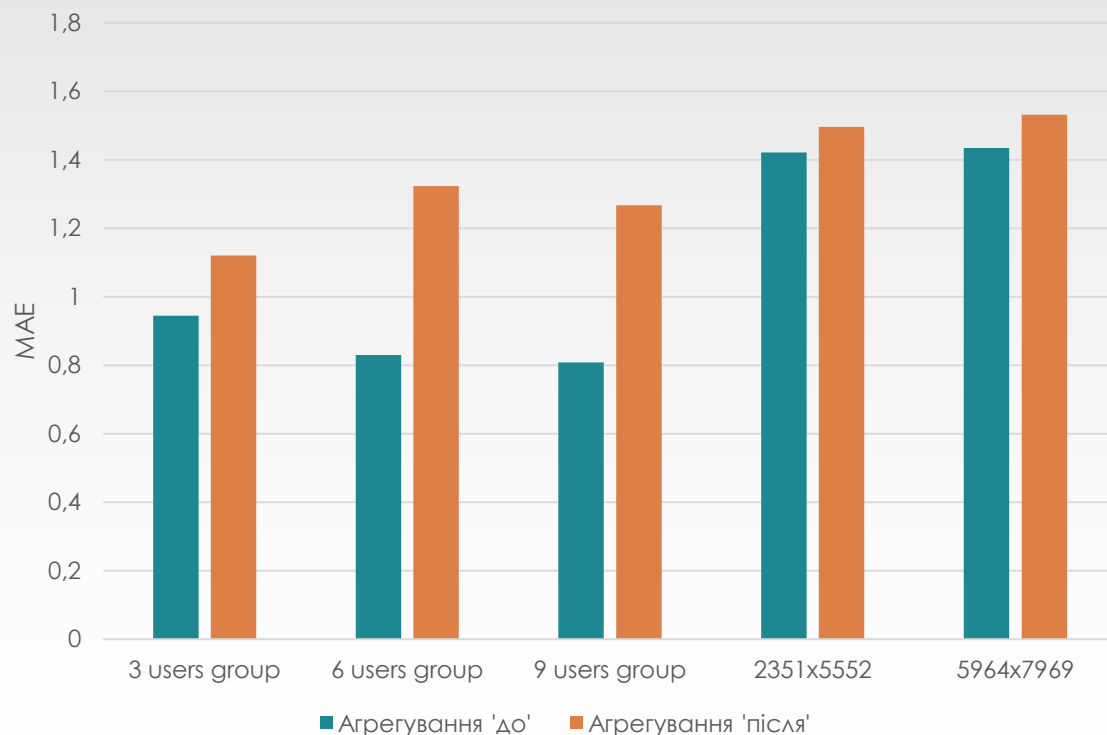


Стратегії суспільного вибору



# Порівняння підходів “до” та “після” за оцінками точності для груп та вибірок різного розміру

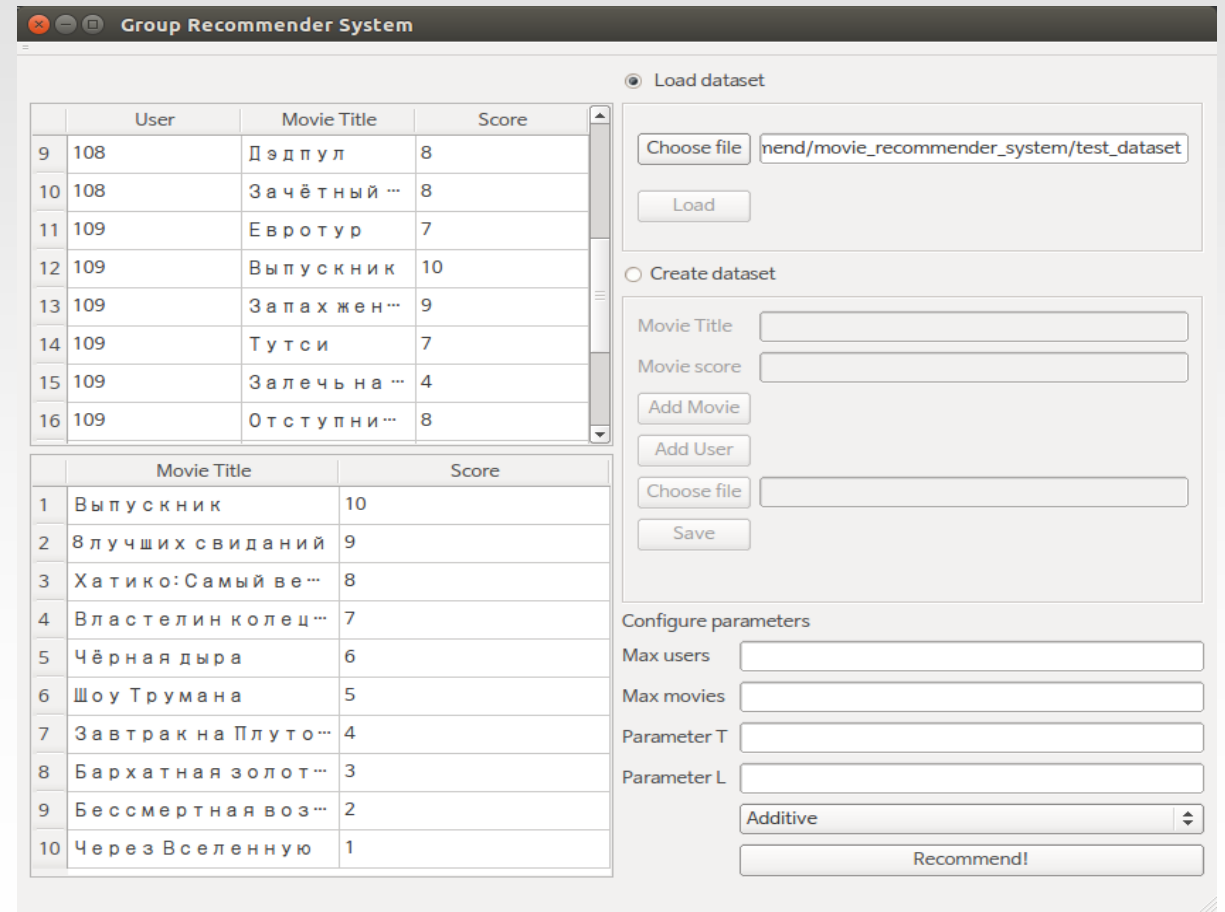
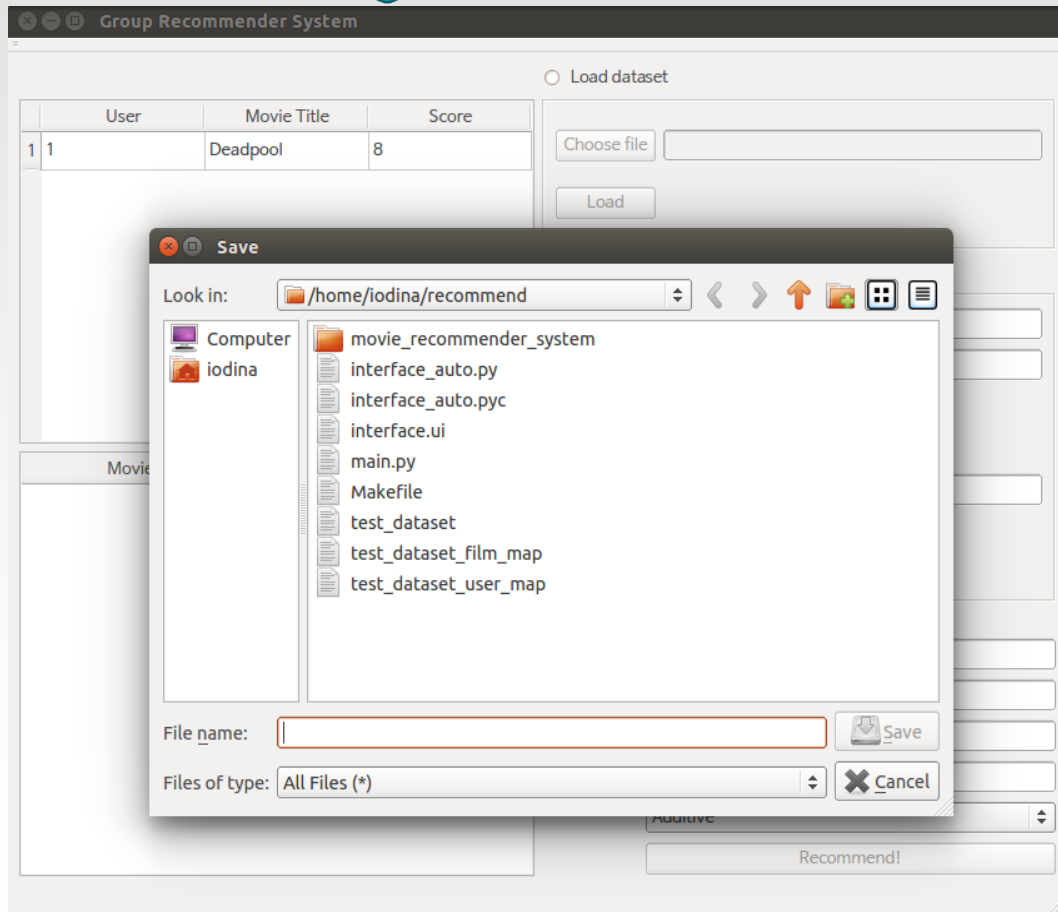
MAE за принципом середнього арифметичного



RMSE за принципом середнього арифметичного



# Интерфейс програми



# ВИСНОВКИ

- Формалізовано методи виявлення спільних вподобань користувачів.
- Запропоновано новий метод оцінювання якості моделей на основі метрики ранжування.
- Проведено оцінювання методів виявлення спільних вподобань користувачів за допомогою метрик точності та власної метрики ранжування.
- Практичним результатом роботи є розробка системи надання групових рекомендацій за індивідуальними вподобаннями групи користувачів.

# Шляхи розвитку

- Оцінювання якості моделі шляхом отримання явного відгуку від користувачів.
- Попередня нормалізація даних.
- Застосування гібридних підходів при формуванні індивідуальних рекомендацій.

**Дякую за увагу**