

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»  
Інститут прикладного системного аналізу  
Кафедра математичних методів системного аналізу

# Рекомендаційна система товарів

## Goods recommendation system

Науковий керівник:  
к.т.н., доцент  
**Дідковська М.В.**

Виконав:  
студент групи КА-51  
**Тарнавський Максим**

# Основні визначення

**Рекомендаційні системи** - це набір технік інформаційного фільтрування, які пропонують користувачам потенційно корисні для них товари.

# Актуальність

Торгівельний гігант **Amazon** заявив, що **35% свого доходу** він отримав завдяки їхнім рекомендаційним системам.

Більше **75% контенту**, який дивились користувачі компанії **Netflix**, був саме запропонований їхньою рекомендаційною системою.

The Google logo, consisting of the word "Google" in its characteristic multi-colored font (blue, red, yellow, green, red).The Facebook logo, featuring the word "facebook" in white lowercase letters on a dark blue rectangular background.The Amazon logo, with the word "amazon" in black lowercase letters and a yellow curved arrow underneath it.The Netflix logo, with the word "NETFLIX" in bold, red, uppercase letters.

# Мета, предмет та об'єкт

**Метою роботи** є розробка рекомендаційної системи для вибору фільмів.

**Об'єктом дослідження** є рекомендаційні системи товарів.

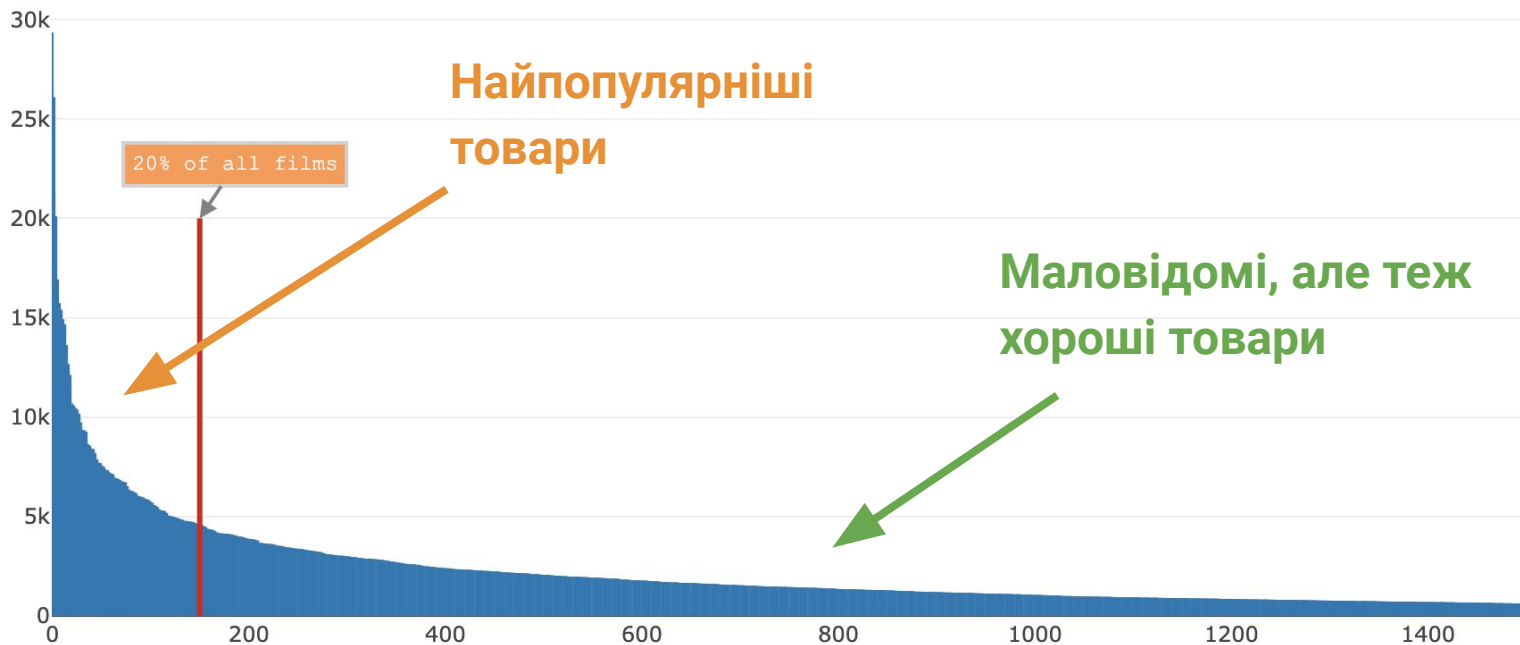
**Предметом дослідження** є методи та алгоритми формування рекомендацій.

# Постановка задачі

- провести дослідження існуючих методів формування рекомендацій
- розробити власний алгоритму формування рекомендацій на основі дворівневої моделі
- провести оцінку якості роботи запропонованого алгоритму
- програмно реалізувати рекомендаційну системи на основі розробленого алгоритму

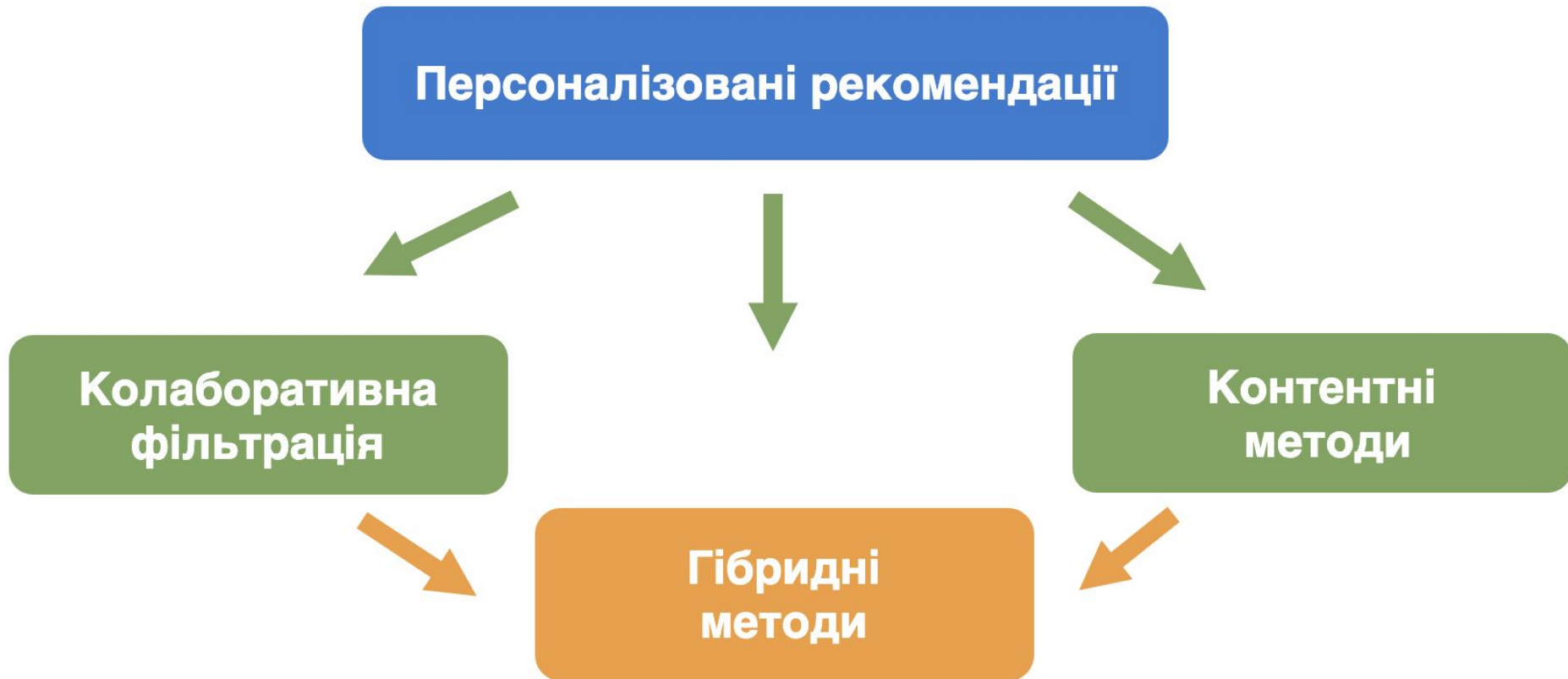
# Існуючі підходи до розв'язку та їхні недоліки

Кількість переглядів

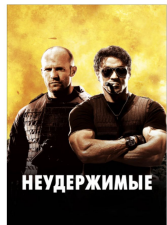


Відсортовані за популярністю товари

# Існуючі підходи до розв'язку та їхні недоліки



# Матриця взаємодій між користувачами та товарами



2

5

3

?



5

2

4

?



5

2

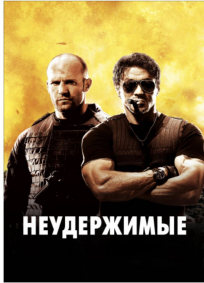
4

?



# Функціональна залежність вподобання користувача

$$f_u : I \rightarrow \mathbb{R}$$

$$f_u \left( \text{НЕУДЕРЖИМЫЕ} \right) = 4.5$$


$$f_u \left( \text{БОЛЬШОЙ СОБАЧИЙ ПОБЕГ} \right) = 2.7$$


# Існуючі підходи до розв'язку та їхні недоліки

## Колаборативна фільтрація

- + Швидкість
- + Нескладні в реалізації
- + Потребують мінімальний об'єм інформації
- Не враховують додаткову інформацію
- Проблема холодного старту

## Контентні методи

- + Більша релевантність рекомендацій
- + Можливість навчатись на підвбірці користувачів
- Значно повільніші
- Ресурсоємні
- Потребують додаткову інформацію

# Існуючі підходи до розв'язку та їхні недоліки

## Гібридні методи

- + Поєднують переваги обох підходів, пропонуючи хороші рекомендації за прийнятний час

# Запропонований алгоритм

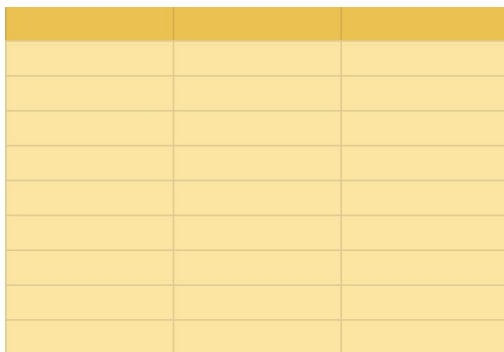
Всі товари



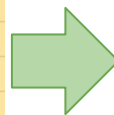
A large blue grid representing all goods. It consists of 13 rows and 3 columns. The top row is a darker blue, while the remaining 12 rows are a lighter blue.



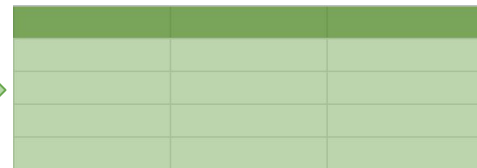
Відбираємо  
товари на 1 рівні



A smaller yellow grid representing goods selected on level 1. It consists of 8 rows and 3 columns. The top row is a darker yellow, while the remaining 7 rows are a lighter yellow.



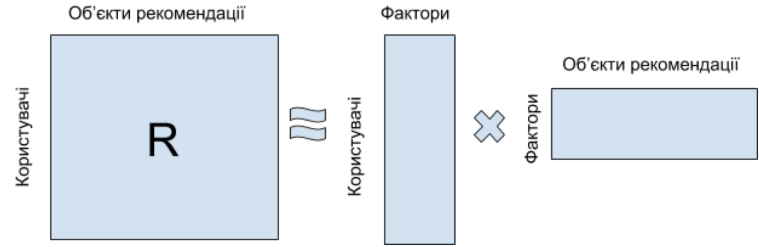
Відбираємо  
товари на 2 рівні



A very small green grid representing goods selected on level 2. It consists of 4 rows and 3 columns. The top row is a darker green, while the remaining 3 rows are a lighter green.

# Модель 1 рівня

## Матрична факторизація



$u \in U$   $i \in I$  - множини користувачів та товарів

$x_u \in R^f$   $y_i \in R^f$  - фактор вектори, які описують користувача  $u$  та товар  $i$

$r_{ui}$  - міра вподобання користувачем  $u$  товару  $i$

$p_{ui}$  - індикатор наявності взаємодії між користувачем  $u$  та товару  $i$

$$c_{ui} = 1 + \alpha r_{ui} \quad \text{Мета: } \hat{p}_{ui} = x_u^T y_i$$

$$\min_{x_*, y_*} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

# Модель 2 рівня

## Градiєнтний бустинг дерев рiшень



**Дерева рiшень** зазвичай представляють собою бiнарнi дерева, в яких в кожнiй внутрiшнiй вершинi записана умова, а в кожному листi дерева - прогнозоване значення

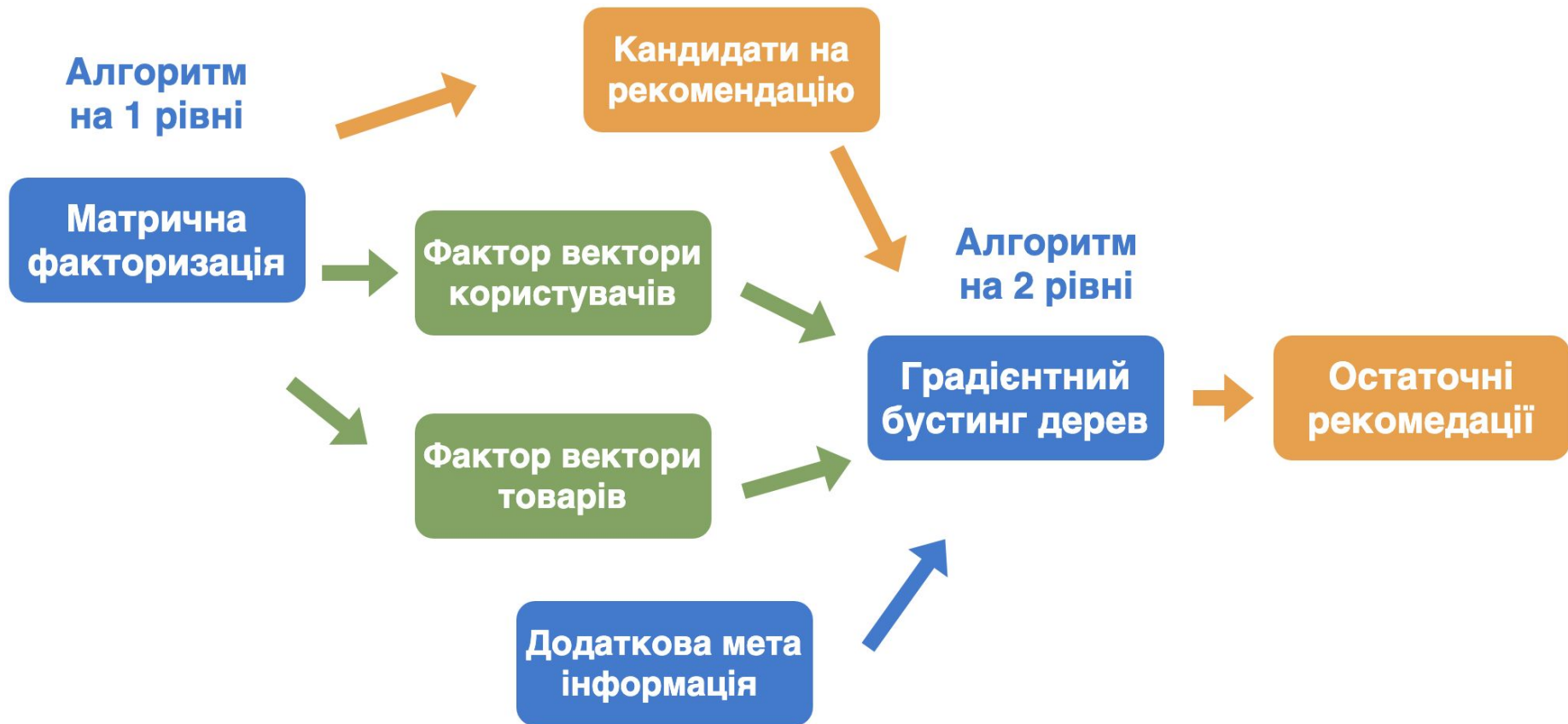
$X_\ell = \{x \in X_m \mid [x^j \leq t]\}$   $X_r = \{x \in X_m \mid [x^j > t]\}$  - подiл умовою на пiдмножини

$Q(X_m, j, t) = \frac{|X_\ell|}{|X_m|} H(X_\ell) + \frac{|X_r|}{|X_m|} H(X_r)$  - умову пiдбирають так, щоб мiнiмiзувати критерiй похибки

$H(X) = \sum_{k=1}^K p_k(1 - p_k)$  - критерiй iнформативностi Джинi

$$p_k = \frac{1}{X} \sum_{i \in X} [y_i = k]$$

# Запропонований алгоритм







# Критерії порівняння

## Online метрики

Бізнес показники, такі як:

- Кількість покупок
- Час проведений в системі
- CTR - відношення числа кліків до числа показів

## Offline метрики

Для прогнозу оцінки:

- RMSE
- MAE

Для оцінки якості рекомендації:

- Precision
- MAP
- NDCG

# Критерії порівняння

$$RMSE = \frac{1}{N} \sum_{ui} \sqrt{(r_{ui} - \hat{r}_{ui})^2} \quad MAE = \frac{1}{N} \sum_{ui} |r_{ui} - \hat{r}_{ui}|$$

$r_{ui}$  - істинне значення оцінки  
 $\hat{r}_{ui}$  - прогнозоване значення оцінки

*Relevance(k)* - індикатор релевантності рекомендованого товару на k-тій позиції, може набувати значення {0,1}

$$Precision@K = \frac{1}{K} \sum_{k=1}^K Relevance(k)$$

$$Average\ Precision@K = \frac{1}{K} \sum_{k=1}^K Relevance(k) \cdot (Precision@k)$$

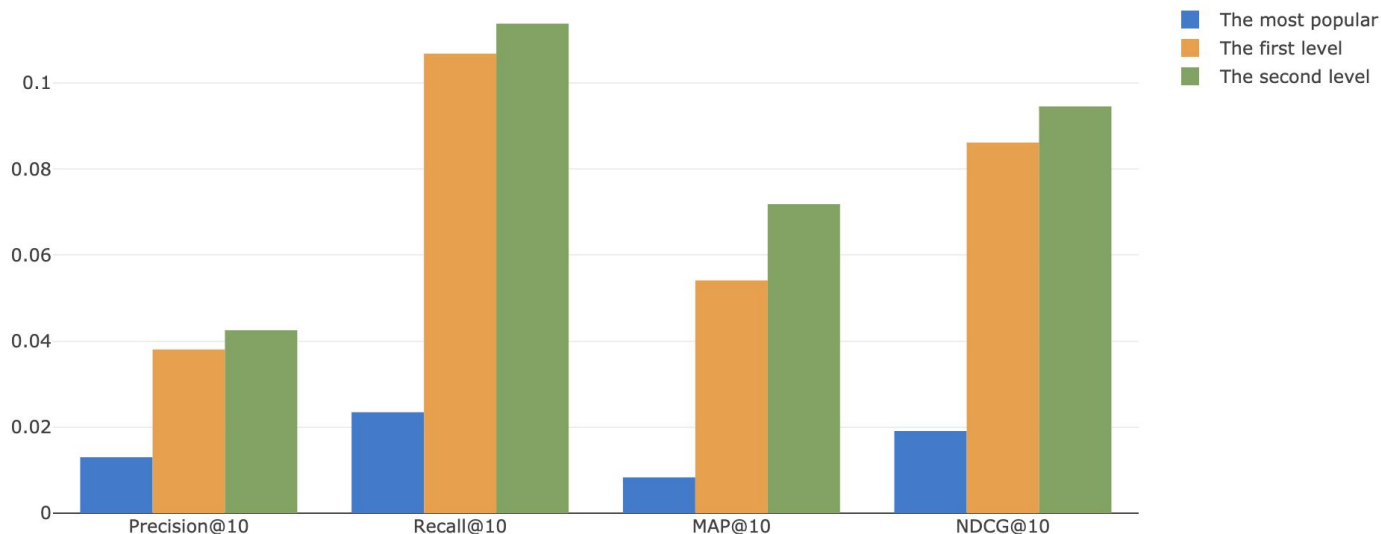
$$MAP@K = \frac{1}{N} \sum_{j=1}^N Average\ Precision@K_j$$

$$DCG@K = \sum_{k=1}^K \frac{Relevance(k)}{\log_2(k+1)}$$

$$NDCG@K = \frac{DCG@K}{\max(DCG@K)}$$

# Порівняльний аналіз практичних результатів

	Precision@10	Recall@10	MAP@10	NDCG@10
The second level	0.0425413	0.113818	0.0718531	0.0945739
The first level	0.038066	0.106846	0.0541142	0.0861713
The most popular	0.0130161	0.0234716	0.00832245	0.0191037



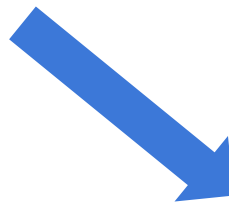
# Порівняльний аналіз практичних результатів

Результат дворівневої моделі

0.07185

Результат 1 місяця в змаганні

0.07107



#	Δpub	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	—	<b>x0x0w1</b>			0.07107	35	2mo
2	—	<b>AfterParty</b>			0.06882	51	2mo
3	—	<b>Netflix</b>			0.06301	39	2mo
4	—	<b>Closer</b>			0.06139	50	2mo
5	—	<b>return_0</b>			0.06035	34	2mo
6	—	<b>zteam</b>			0.05963	25	2mo

Моя команда на момент змагання

# Демонстрація роботи продукту

## MEGOGO

### Recommended films



	Metrics
Prostori 10	0.3
MEGOGO	1.0
MAPS 10	0.7667
INDIGO 10	0.301

Select good recommendations and

Get metrics

# Висновки



CIKLUM  
EMPOWERING COLLABORATION

- запропоновано алгоритм дворівневої моделі для здійснення рекомендацій, який поєднує собі матричну факторизацію та градієнтний бустинг дерев
- розроблений алгоритм отримав значення метрики MAP@10 **0.07185** в змаганні Meqogo, що перевищує найкращий результат запропонований на момент змагання
- даний алгоритм впроваджено в компанії **Ciklum**

# Шляхи подальшого розвитку

Варто розглянути :

- DSSM нейронні мережі
- Тензорна факторизація матриці

**Дякую за увагу !**