

# АНАЛІЗ ТА ПОРІВНЯННЯ ЕФЕКТИВНОСТІ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

Виконала: Зінченко А. Р.

Науковий керівник: Салата К. В.

Національний університет “Києво-  
Могилянська академія”

# МЕТА РОБОТИ

Провести аналіз та порівняння ефективності класичних і сучасних методів рекомендаційних систем (зокрема, *item-based* та *user-based* колаборативної фільтрації, SVD та NCF) з метою визначення їх переваг, недоліків і придатності до застосування в умовах обмежених даних.

# ЗАВДАННЯ РОБОТИ


- Ознайомитися з теоретичними основами рекомендаційних систем й дослідити історію розвитку та сучасні тенденції.
- Розглянути принципи роботи `user-based`, `item-based`, `SVD` та `NCF` методів.
- Реалізувати обрані підходи на основі відкритого датасету (`MovieLens`), провести оцінку ефективності та порівняти результати

# ВИКОРИСТАНИЙ ДАТАСЕТ

## MovieLens Latest Small Dataset

- Містить 100 836 оцінок від 610 користувачів на 9 742 фільми.
- Складається з таблиць: `ratings.csv`, `movies.csv`, `tags.csv`, `links.csv`.
- Має оптимальний розмір для обробки за умов обмежених ресурсів
- Широко використовується у дослідженнях рекомендаційних систем

# ВИБІР МЕТОДІВ

- User-based та Item-based CF – прості й інтерпретовані базові методи (memory-based), ефективні для невеликих систем.
- SVD – популярний model-based підхід, який враховує приховані зв'язки між користувачами та об'єктами.
-  NCF – сучасний нейромережевий метод, здатний моделювати складні залежності, перспективний для масштабних систем.

# ВИБІР МЕТРИК ПОРІВНЯННЯ

- MAE, RMSE – оцінка точності прогнозів (менші значення = краща відповідність).
- Coverage – показує, скільки різних об'єктів система здатна рекомендувати.
- Diversity – оцінює різноманітність списку рекомендацій.
- Novelty – вимірює здатність системи рекомендувати менш відомі, але потенційно цікаві об'єкти.

# ВИБІР ТЕХНОЛОГІЙ

Використані бібліотеки:

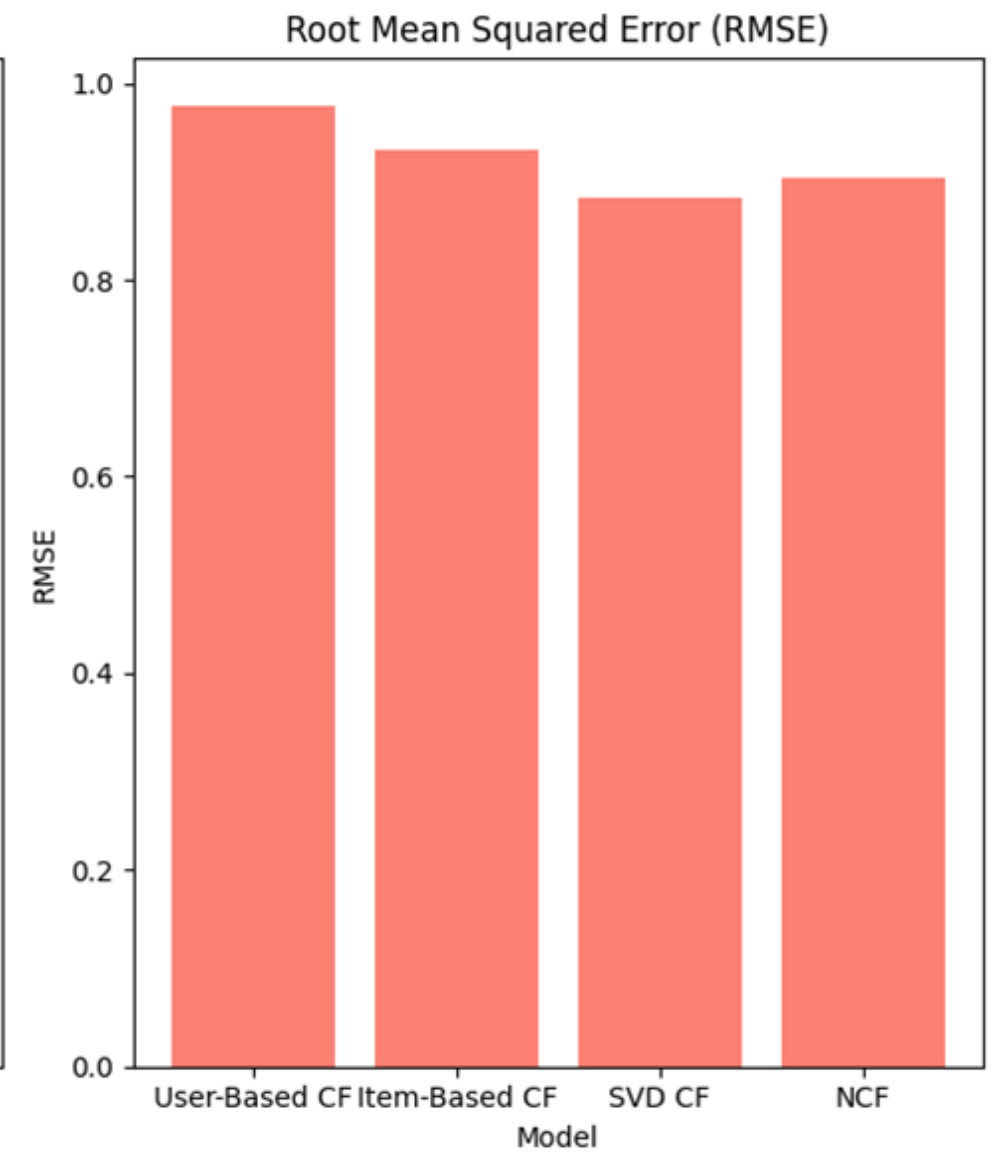
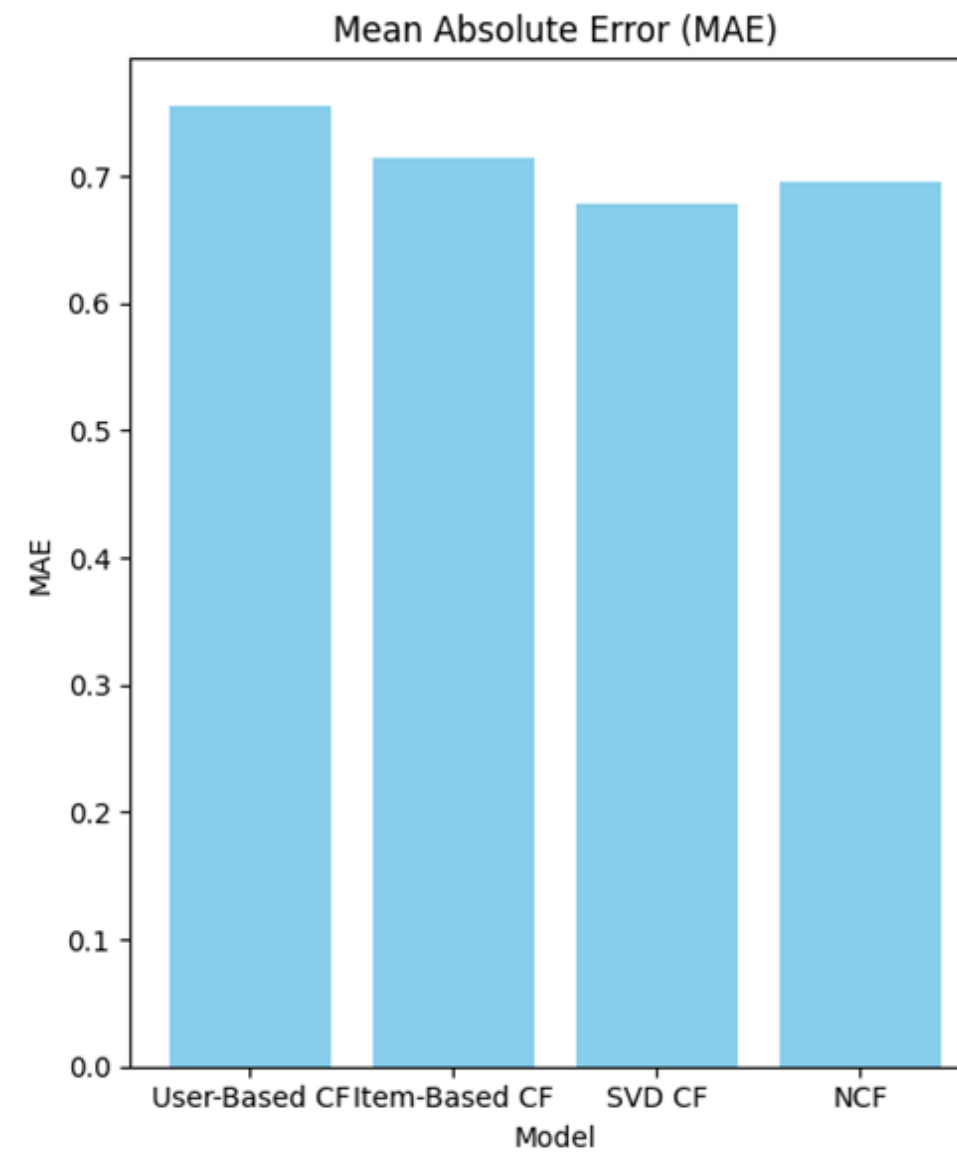
- `pandas`
- `numpy`
- `scikit-learn`
- `surprise`
- `tensorflow / pytorch`
- `matplotlib`



# РЕЗУЛЬТАТИ

## Метрики точності:

	MAE	RMSE
User-Based CF	0,7544	0,9762
Item-Based CF	0,7135	0,9322
SVD CF	0,6748	0,8790
NCF	0,6871	0,9007



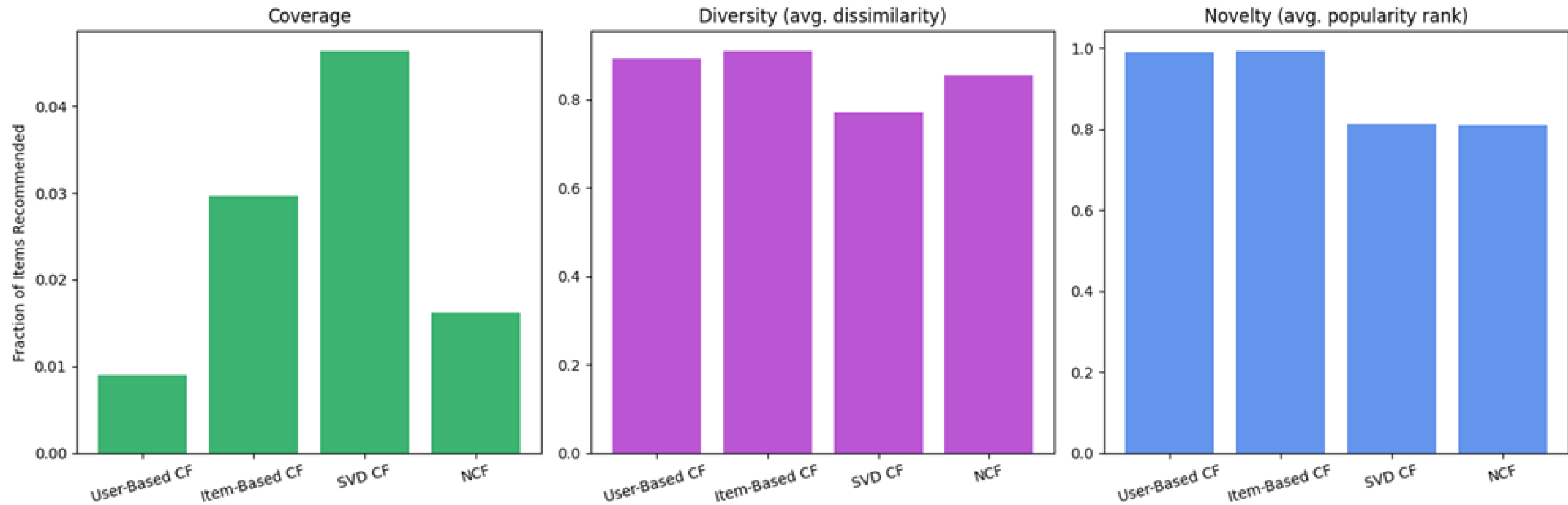
# РЕЗУЛЬТАТИ

**Оцінка покриття, різноманіття і новизни:**

	Coverage	Diversity	Novelty
User-Based CF	0,009	0,893	0,990
Item-Based CF	0,030	0,909	0,992
SVD CF	0,046	0,770	0,811
NCF	0,016	0,854	0,810



# РЕЗУЛЬТАТИ



# ВИСНОВКИ

- Item-based CF демонструє найкращий баланс між точністю, новизною та різноманіттям.
- SVD забезпечує найвищу точність і coverage, але менш інноваційні рекомендації.
- User-based CF добре працює для персоналізованих нішевих рекомендацій, але має слабке охоплення.
- NCF має потенціал, але потребує більше даних та тонкого налаштування.

ДЯКУЮ ЗА  
УВАГУ!