



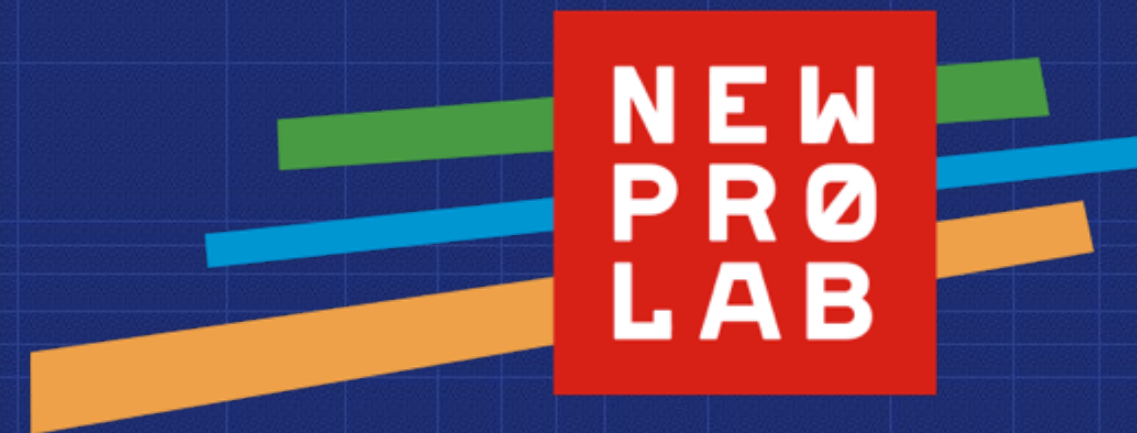
NEW
PRO
LAB

Рекомендательные системы

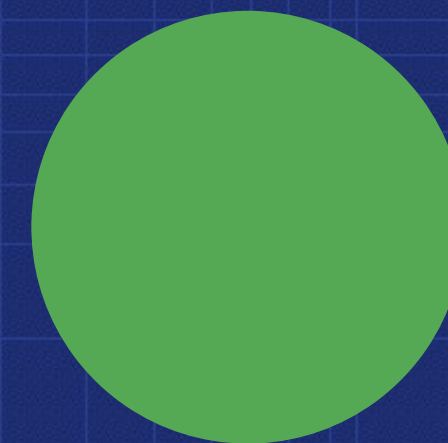
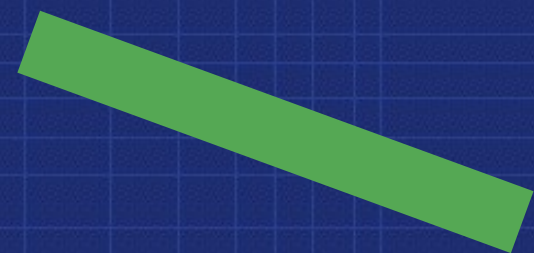
Гриша Сапунов

NEWPROLAB.COM

The background is a dark blue grid. It features various geometric shapes: a large orange circle in the top right, a blue circle in the top right, a small orange circle in the center, a blue circle in the bottom left, a green square in the bottom right, a red triangle in the bottom right, a blue square in the middle right, and several orange and blue lines of different lengths and orientations scattered across the slide.

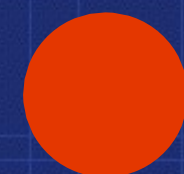
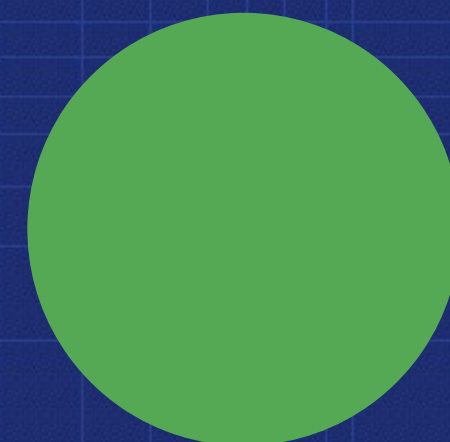
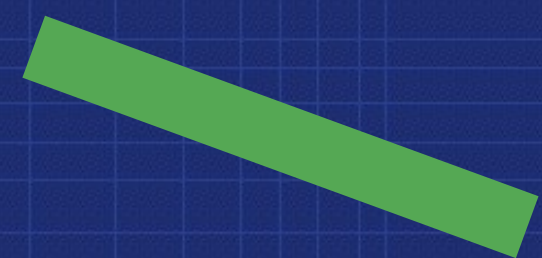


Оценка качества рекомендательных систем





offline эксперименты



Experimental protocols



- hidden data
- cross-validation

Experimental protocols

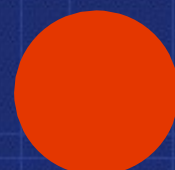
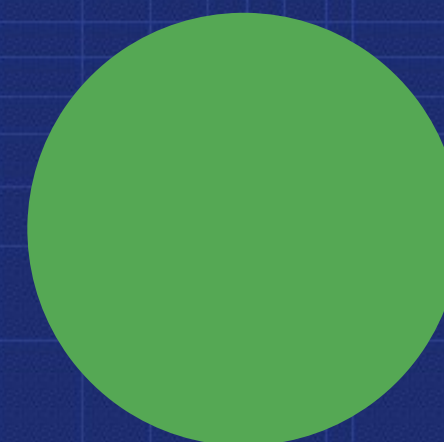
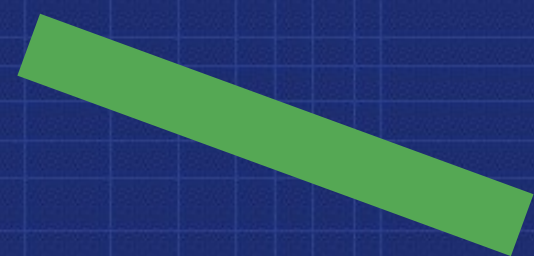
- разбить данные на k блоков
- обучаться на $k-1$
- тестировать на 1

Experimental protocols

- На уровне пользователя для тестовой выборки
 - скрыть случайным образом $x\%$ (x) рейтингов
 - скрыть рейтинги с некоторого момента времени

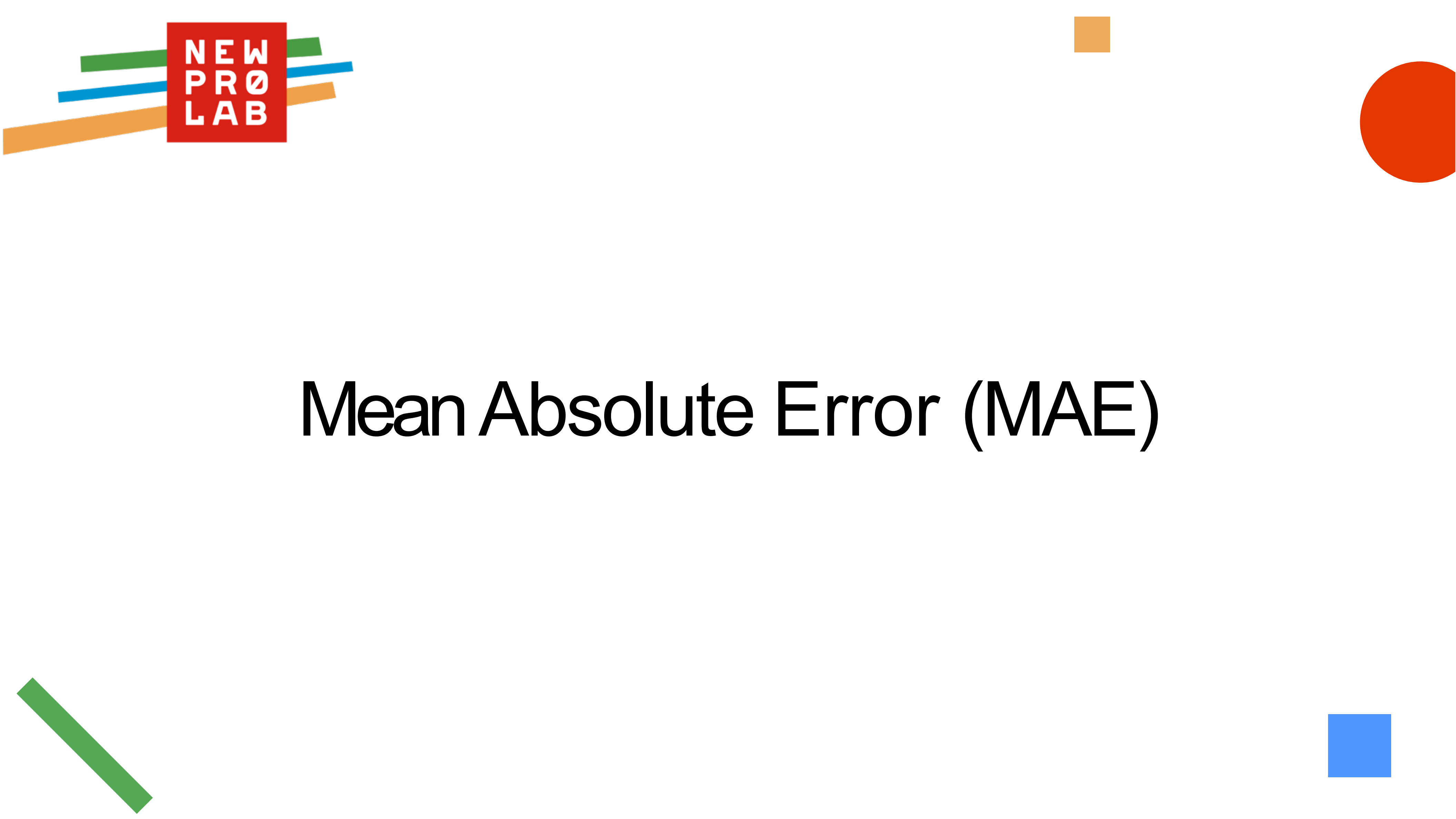


метрики точности предсказания



Точность предсказания

- MAE (Mean Absolute Error)
- MSE (Mean Squared Error)
- RMSE (Root Mean Squared Error)



NEW
PRO
LAB

Mean Absolute Error (MAE)

предсказание

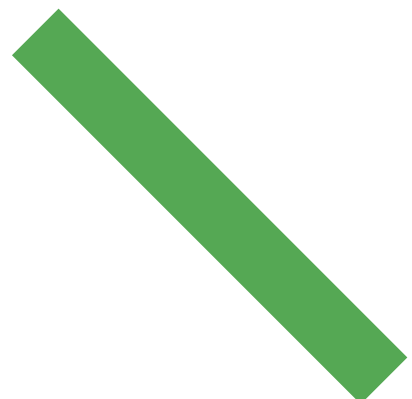
$$MAE = \frac{1}{|K|} \sum_{i \in K, u} |\hat{r}_{ui} - r_{ui}|$$

hidden ratings

- K - множество всех скрытых рейтингов (hidden ratings)
- $|K|$ - количество элементов множества K
- \hat{r}_{ui} - предсказание для товара i и пользователя u
- r_{ui} - истинный рейтинг для товара i и пользователя u



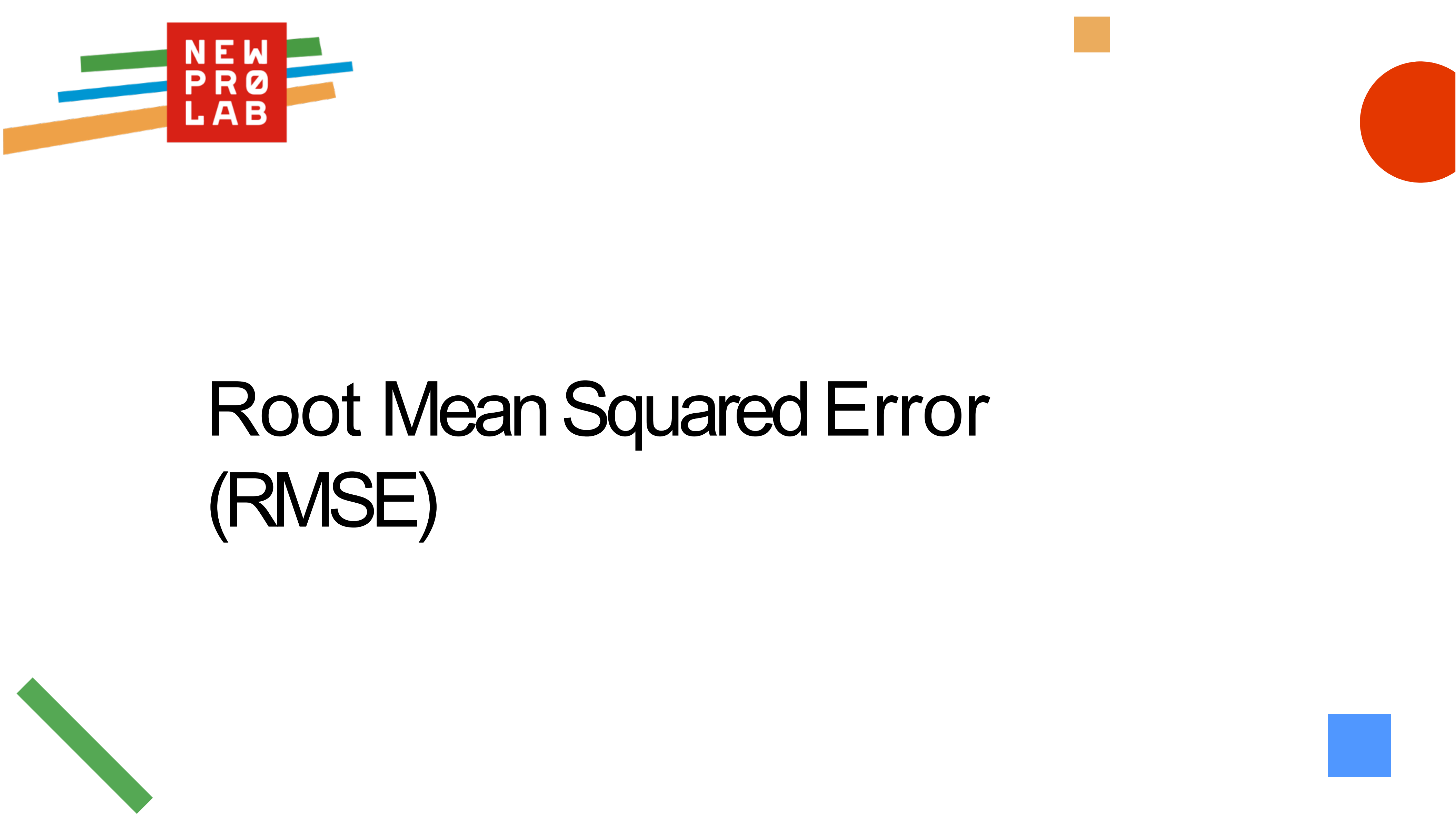
Mean Squared Error (MSE)



предсказание

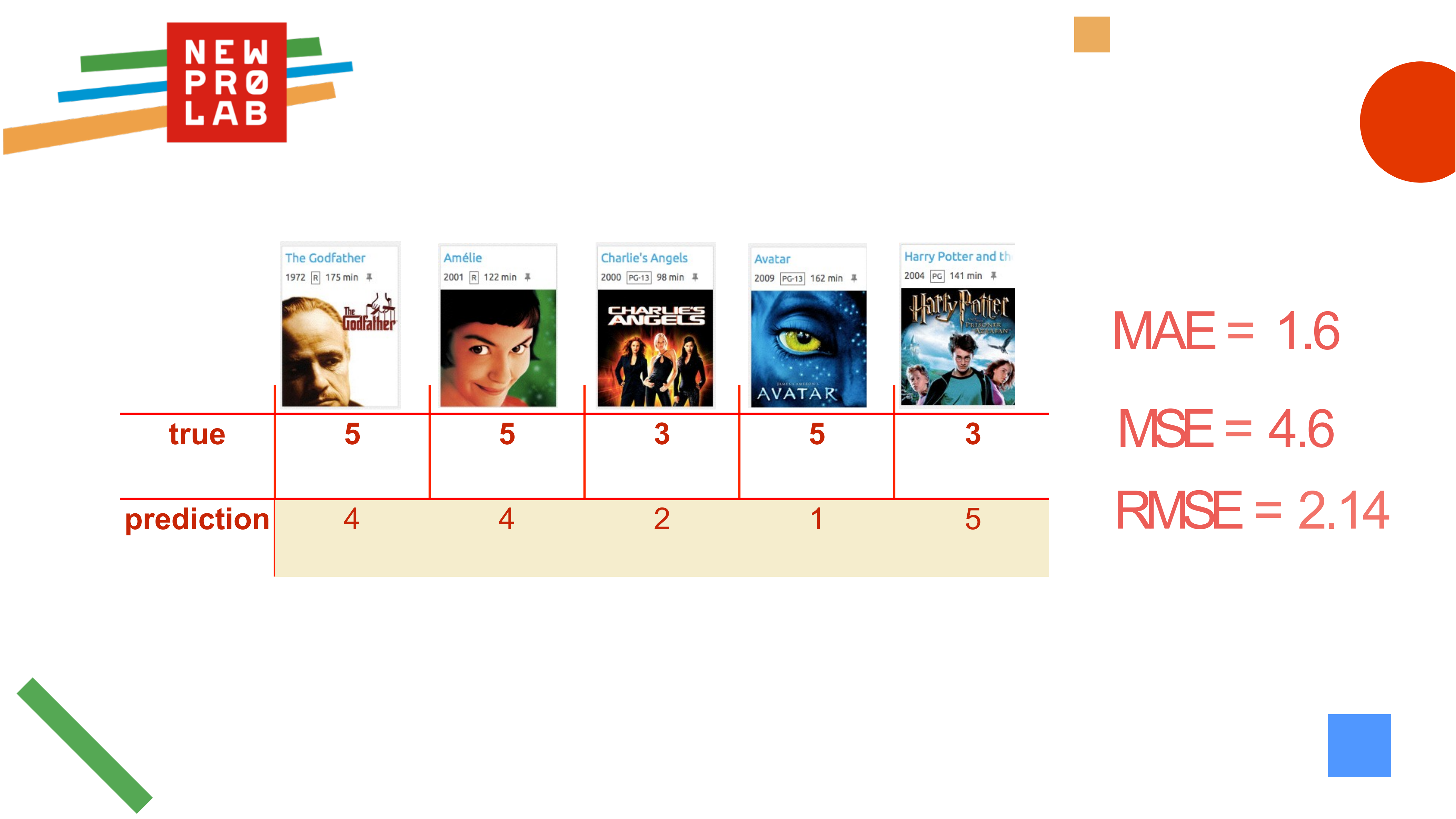
$$MSE = \frac{1}{|K|} \sum_{i \in K, u} \left(\hat{r}_{ui} - r_{ui} \right)^2$$

hidden ratings



Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{|K|} \sum_{u, i \in K} \left(\hat{r}_{ui} - r_{ui} \right)^2}$$

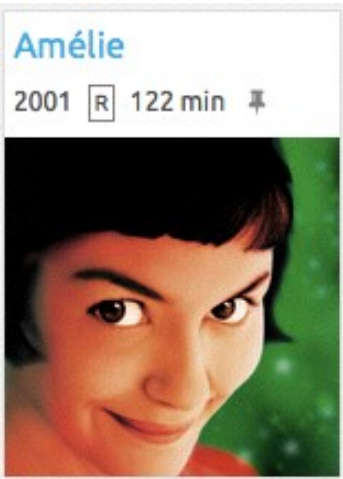
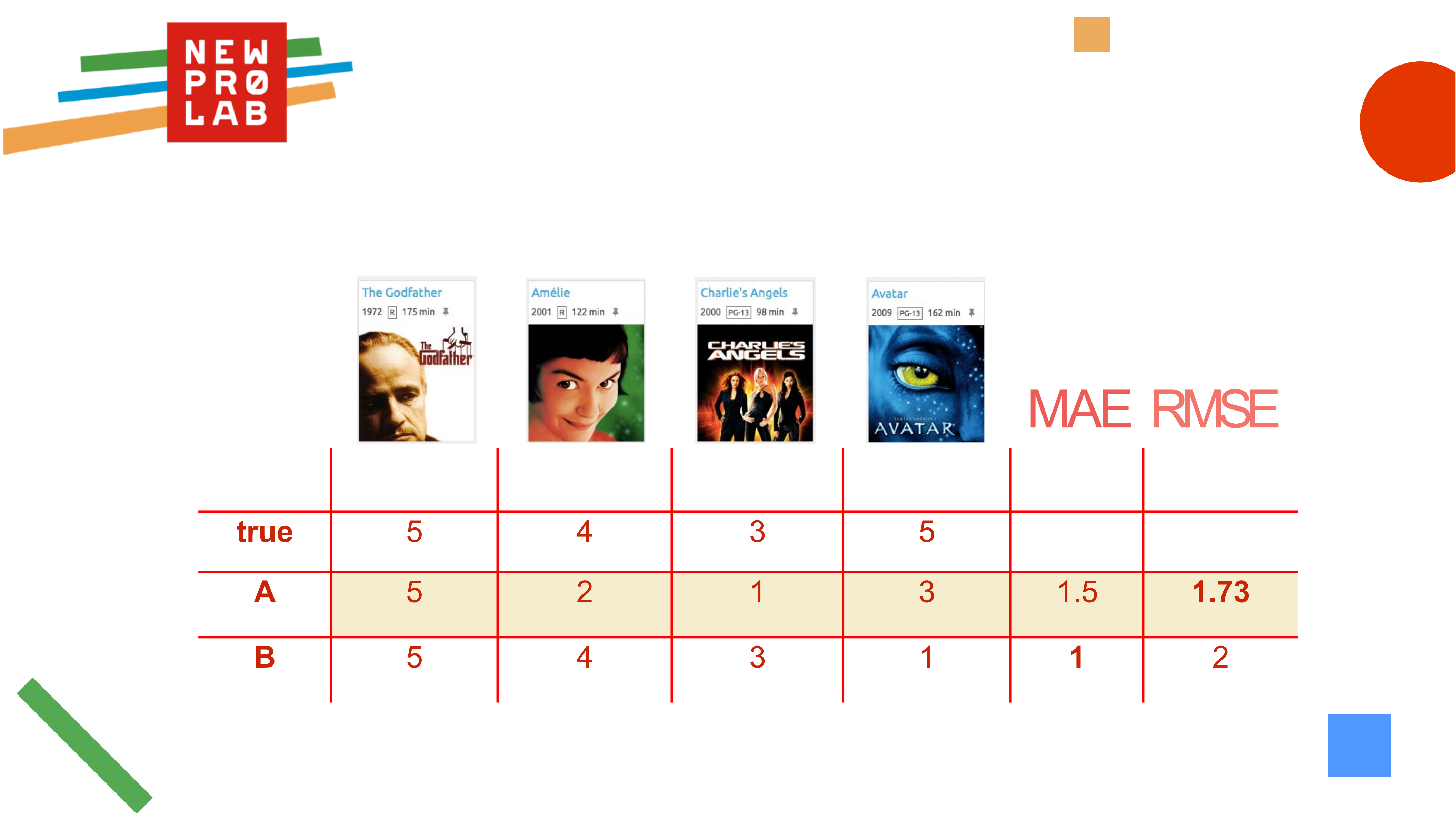


	 <p>The Godfather 1972 [R] 175 min</p>	 <p>Amélie 2001 [R] 122 min</p>	 <p>Charlie's Angels 2000 [PG-13] 98 min</p>	 <p>Avatar 2009 [PG-13] 162 min</p>	 <p>Harry Potter and the Prisoner of Azkaban 2004 [PG] 141 min</p>
true	5	5	3	5	3
prediction	4	4	2	1	5

$$MAE = 1.6$$

$$MSE = 4.6$$

$$RMSE = 2.14$$



MAE RMSE

	The Godfather	Amélie	Charlie's Angels	Avatar		
true	5	4	3	5		
A	5	2	1	3	1.5	1.73
B	5	4	3	1	1	2

Как агрегировать

- по всем рейтингам сразу?
- по пользователям?

Как агрегировать

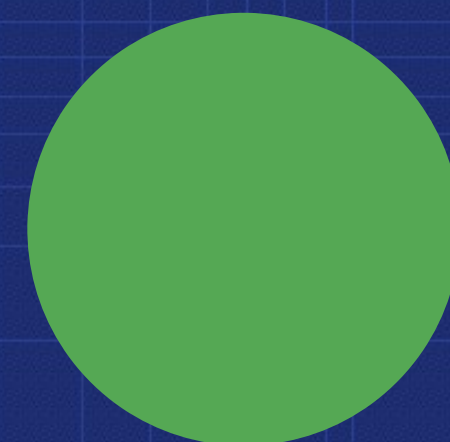
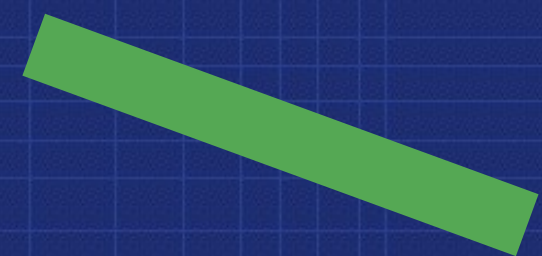
$RMSE_{ratings}$? $RMSE_{users}$

Как сравнить два алгоритма?

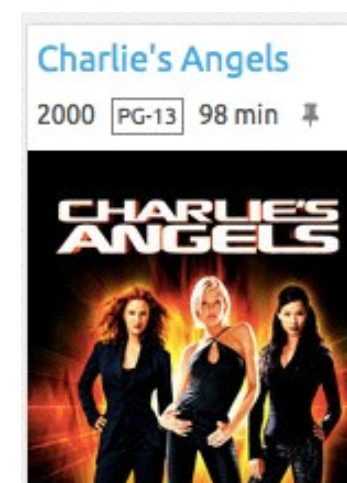
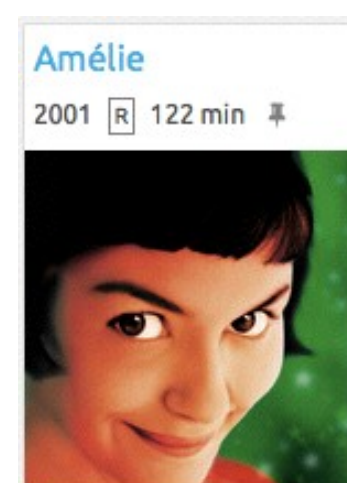
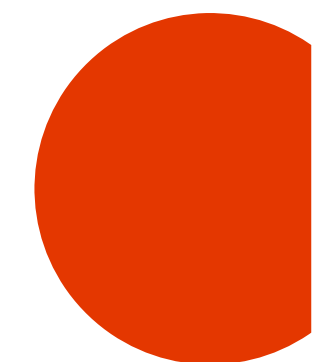
- разница в шкале рейтингов
- одни и те же данные
- что если покрытие разное?



decision-support метрики



- принятие «хороших» решений
- ошибки (errors)



true	5	5	1	5	4
prediction	2	4	4	4	5

$3.5^* - 5^* = \text{good}$

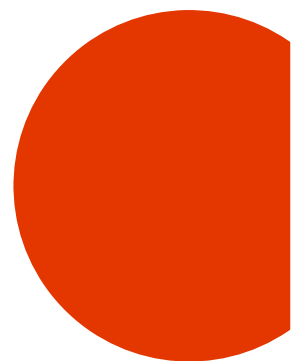
$0.5^* - 2.5^* = \text{bad}$

errors = 2

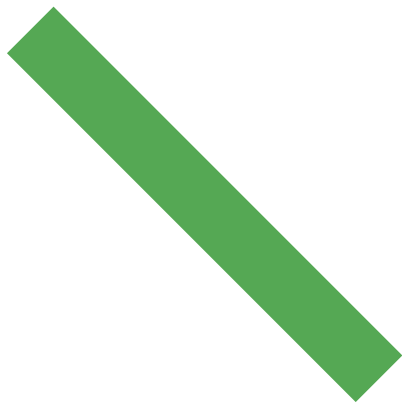




Precision and Recall



	Recommended	Not Recommended	
Preferred	True-Positive (tp)	False-Negative (fn)	$\#tp + \#fn$
Not Preferred	False-Positive (fp)	True-Negative (tn)	$\#fp + \#tn$
	$\#tp + \#fp$		



Precision

$$P = \frac{\#tp}{\#tp + \#fp}$$

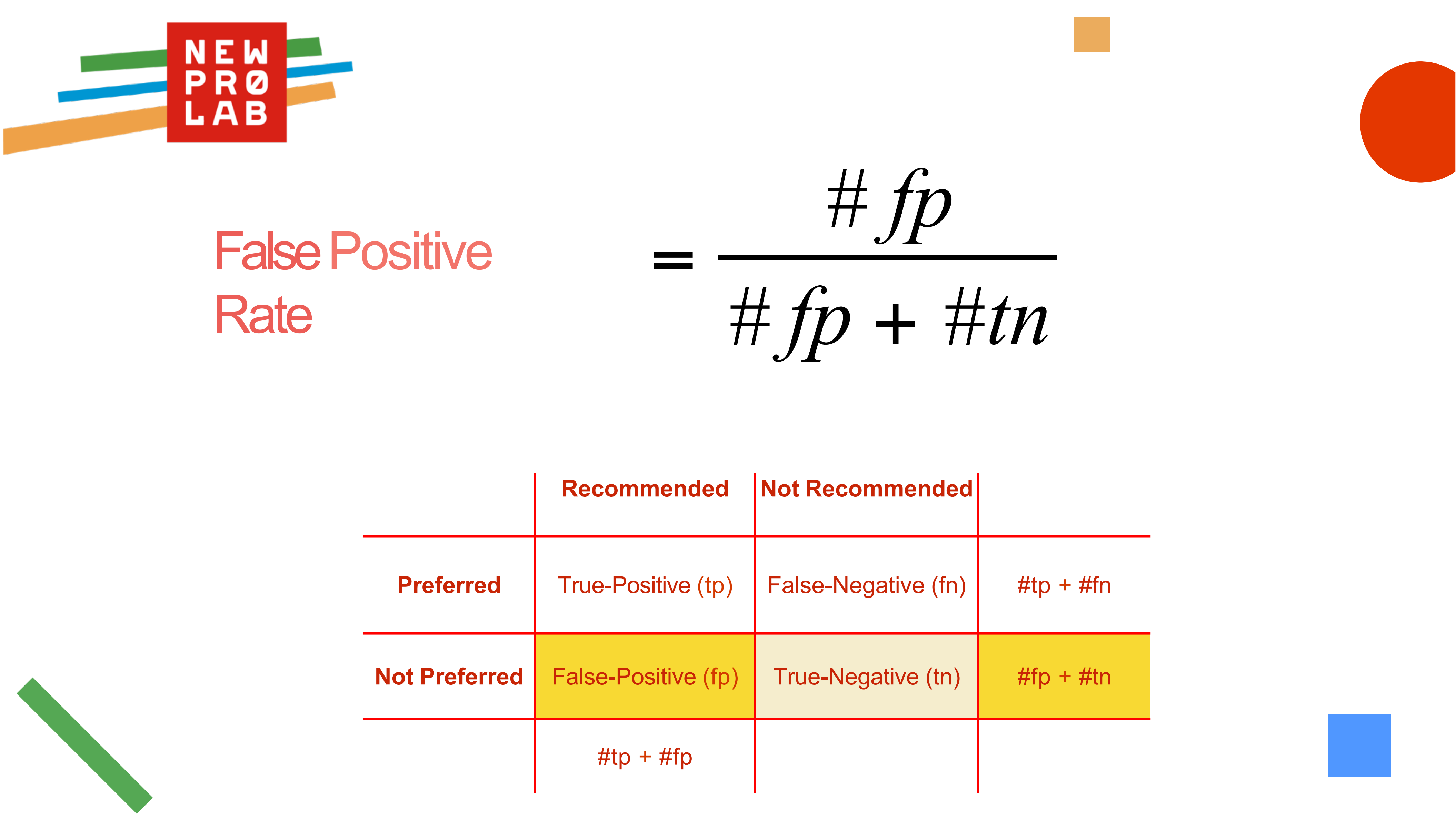
	Recommended	Not Recommended	
Preferred	True-Positive (tp)	False-Negative (fn)	#tp + #fn
Not Preferred	False-Positive (fp)	True-Negative (tn)	
	#tp + #fp		

Recall

True Positive Rate

$$R = \frac{\#tp}{\#tp + \#fn}$$

	Recommended	Not Recommended	
Preferred	True-Positive (tp)	False-Negative (fn)	#tp + #fn
Not Preferred	False-Positive (fp)	True-Negative (tn)	
	#tp + #fp		

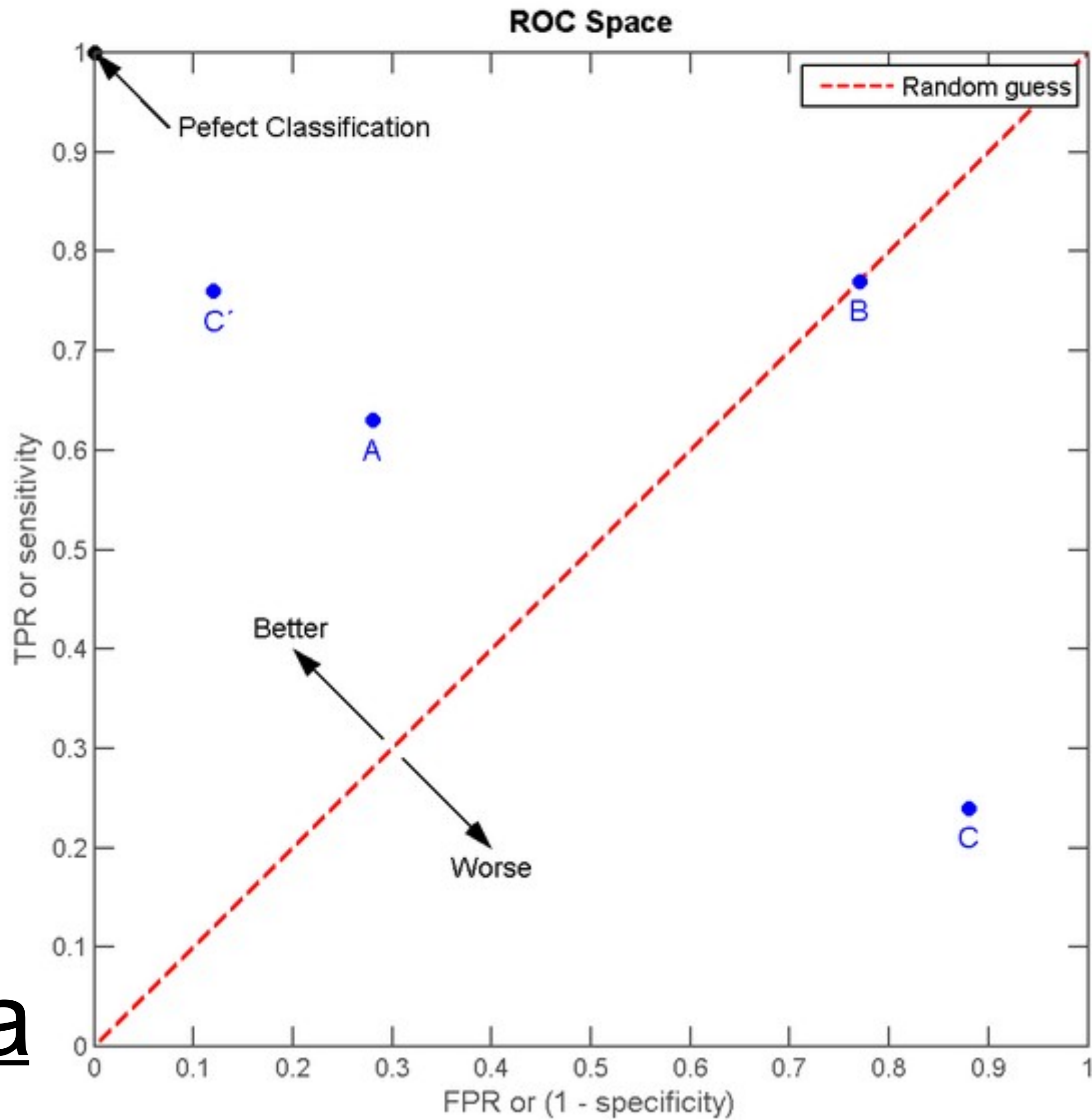


False Positive
Rate

$$= \frac{\#fp}{\#fp + \#tn}$$

	Recommended	Not Recommended	
Preferred	True-Positive (tp)	False-Negative (fn)	#tp + #fn
Not Preferred	False-Positive (fp)	True-Negative (tn)	#fp + #tn
	#tp + #fp		

- precision/recall curve
- true positive rate/false positive rate (ROC – Receiver Operating Characteristic)



Что такое Recall в RecSys?

- прогнозировать топ по всем данным,
засчитывать только оцененные раньше
- ограничиться только оцененными товарами
- попросить людей оценить

Table 7.4. *Example user ratings.*

Row	UserID	MovieID	Rating
1	234	110	5
2	234	151	5
3	234	260	3
4	234	376	5
5	234	539	4 ^a
6	234	590	5
7	234	649	1
8	234	719	5 ^a
9	234	734	3
10	234	736	2

^a Randomly selected ratings for testing.

$$recset_{234} = \{(912, 4.8), (47, 4.5), (263, 4.4), (\mathbf{539}, \mathbf{4.1}), (348, 4), \dots, (\mathbf{719}, \mathbf{3.8})\}$$

комбинированная
метрика P & R

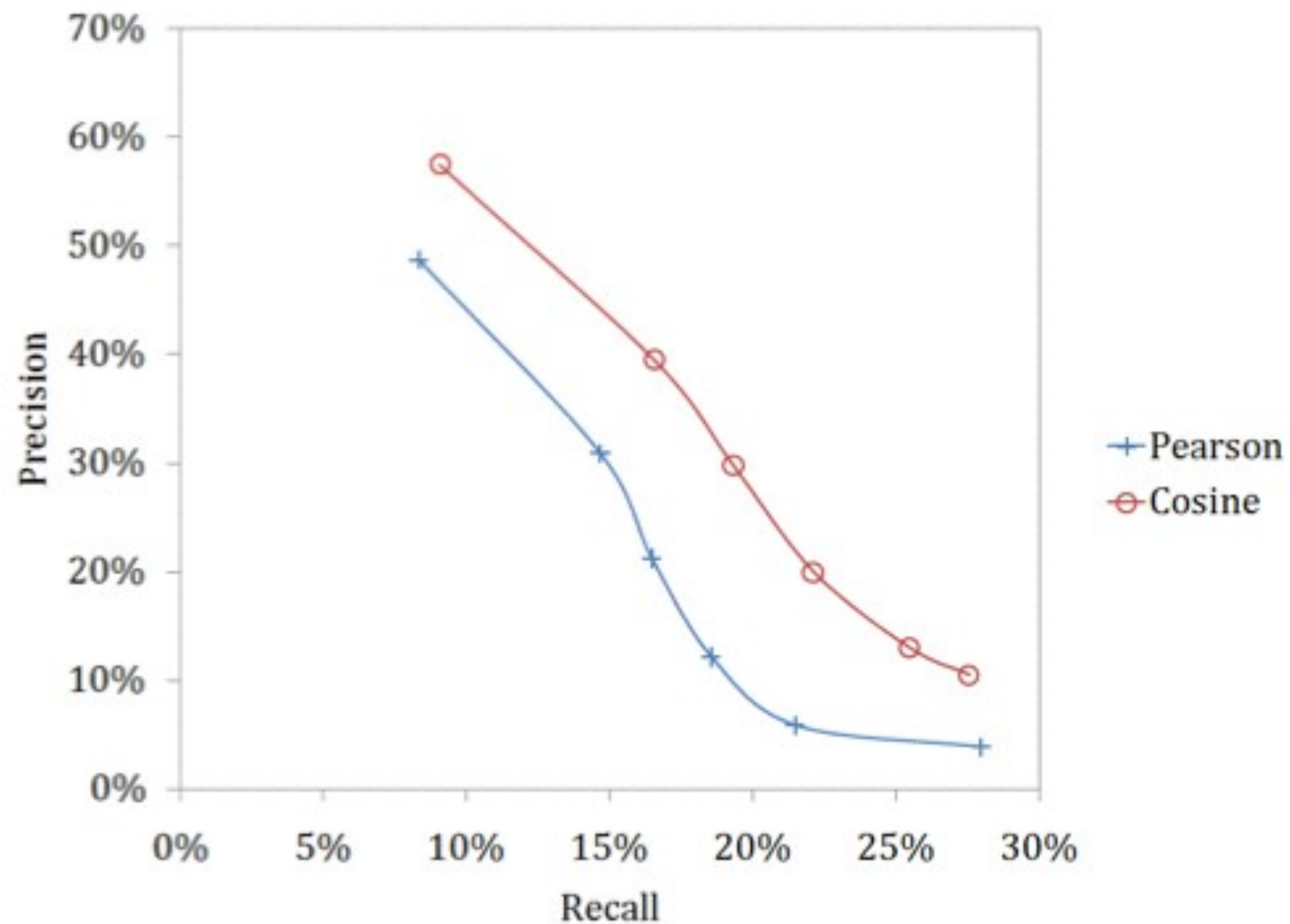
$$F1 = \frac{2 \cdot P \cdot R}{P + R}$$

- $P@k$
- $R@k$
- усреднение по пользователям
- $MAP@k$ (Mean Average Precision at K)

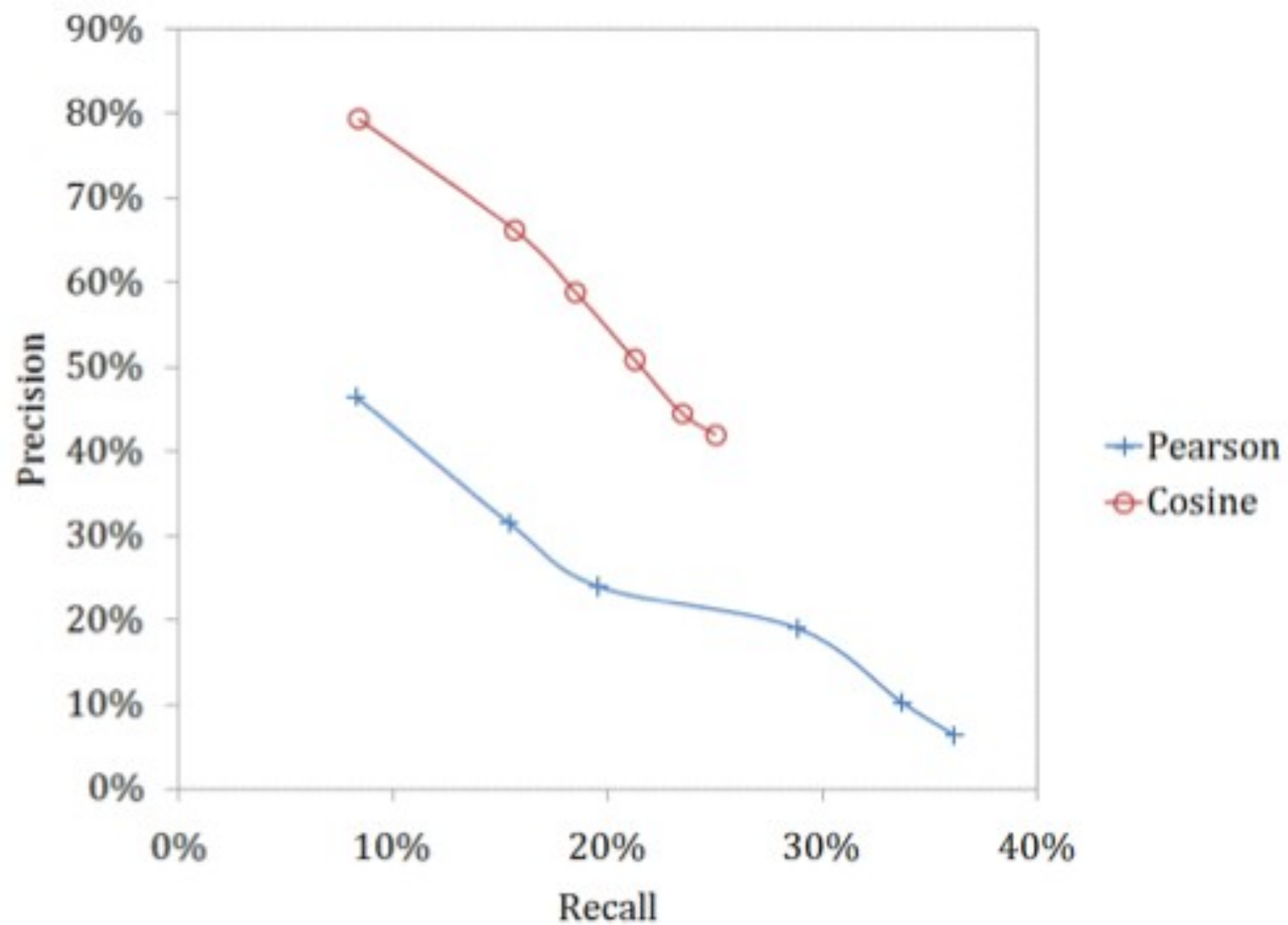
RMSE

	Netflix	BookCrossing
Pearson	1.07	3.58
Cosine	1.90	4.5

1, 3, 5, 10, 25 и 50
рекомендаций



(a) Belgian retailer recommendations



(b) News click stream recommendations

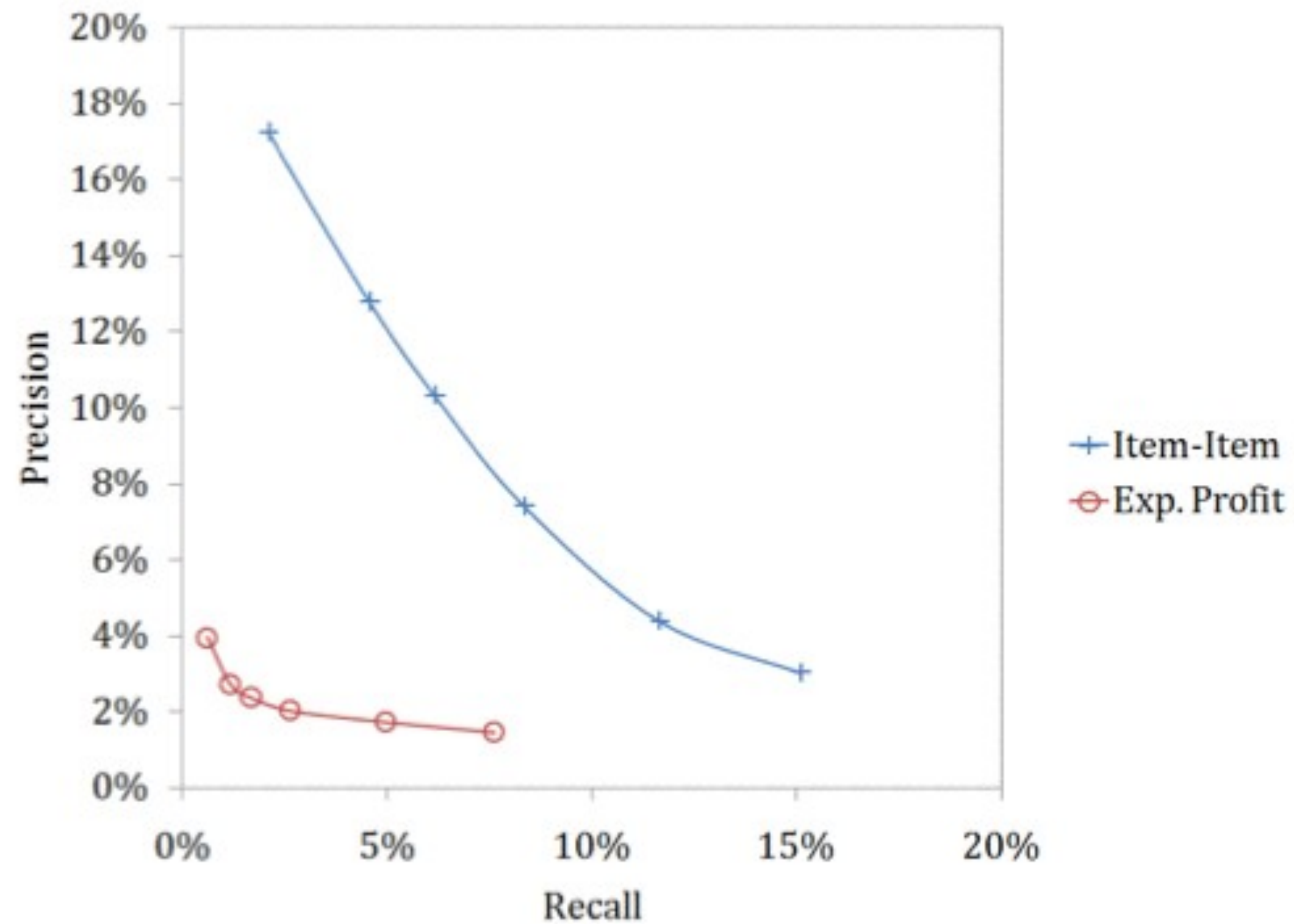


Figure 3: Comparing recommendations generated by the item-item recommender and the expected profit recommender on the Ta-Feng data set.

Базовая метрика - Precision (точность).

Измеряем долю позитива в рекомендованном списке длины N_{rec}



$$Precision = \frac{| \text{Угаданные позитивные объекты} |}{| \text{Рекомендованные объекты} |} = \frac{N_{rate>5}}{N_{rec}}$$

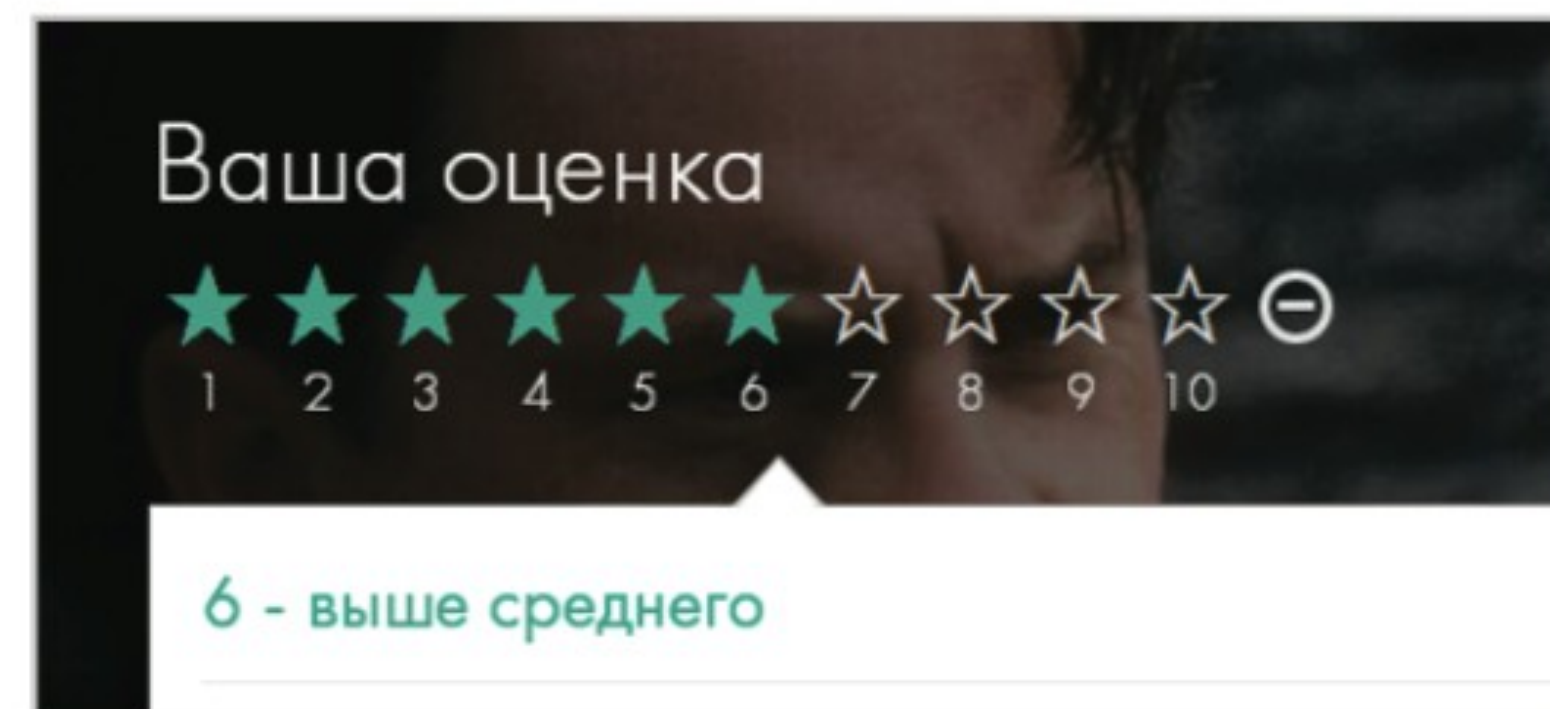
Обобщаем формулу на всех пользователей сервиса

$$Precision = \frac{\sum_{\text{все пользователи}} | \text{Угаданные позитивные объекты} |}{\sum_{\text{все пользователи}} | \text{Рекомендованные объекты} |} = \frac{\sum_{\text{все пользователи}} N_{rate>5}}{N_{rec} | \text{Все пользователи} |}$$

Учёт баллов в Precision

Формулу для Precision можно усовершенствовать и учитывать не только факт того, что фильм понравился пользователю, но и то насколько сильно он понравился.

В таком виде метрика Precision реагирует не только на концентрацию позитивных оценок в рекомендованном списке, но и на их качество.



$$Precision^* = \frac{\sum_{rate > 5} (rate - 5)}{N_{rec}}$$

Дисконтированная метрика

Если нам известно, какая доля пользователей w_N останавливается на N-ом элементе (далее не просматривает), то полная усредненная удовлетворенность всей аудитории должна представлять собой взвешенную таким образом сумму значений Precision:

$$AUC = \sum_N Pr_N w_N = \sum_N \frac{\sum_{rate > 5} (rate - 5)}{N \cdot |Users|} w_N$$