



Система персональных рекомендаций Гибридная модель

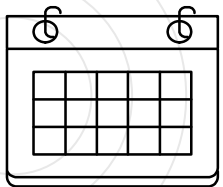


Всем привет!



Меня зовут Дарья!

Этапы проекта



1 Модель с **ALS**

2 Битва с **RankFM**

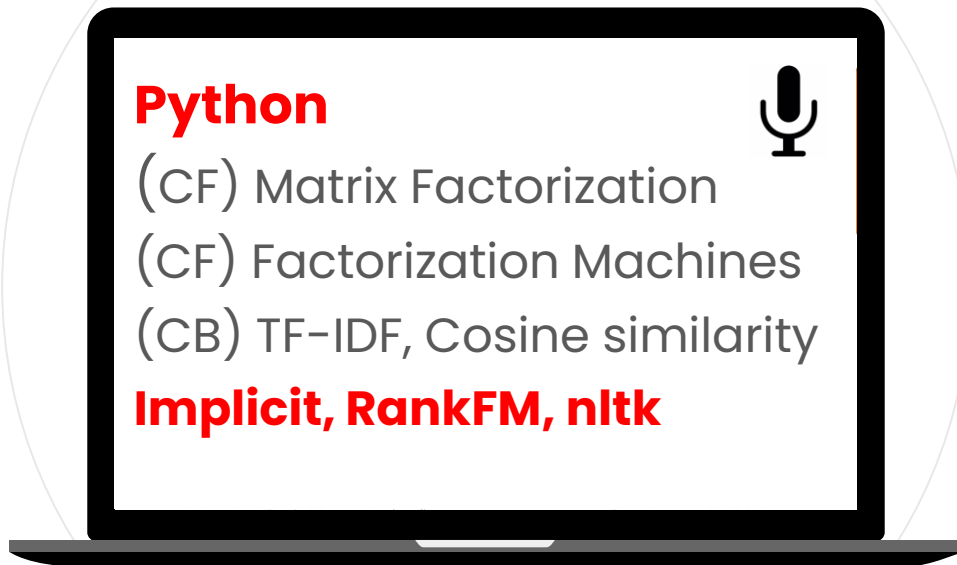
3 Возможности **CF**

4 Интерпретация модели

5 Рекомендации с **TF-IDF**

6 Гибридное решение

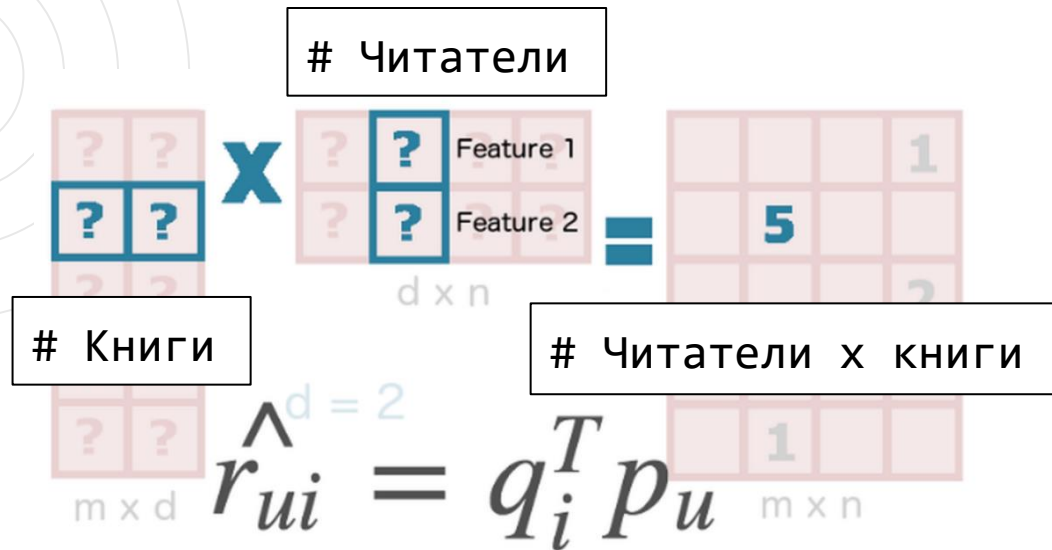
Технологии



CF – Collaborative Filtering

CB – Content Based Filtering

Collaborative Filtering (CF)



*Matrix
Factorization*

Alternating
Least Squares

#Implicit



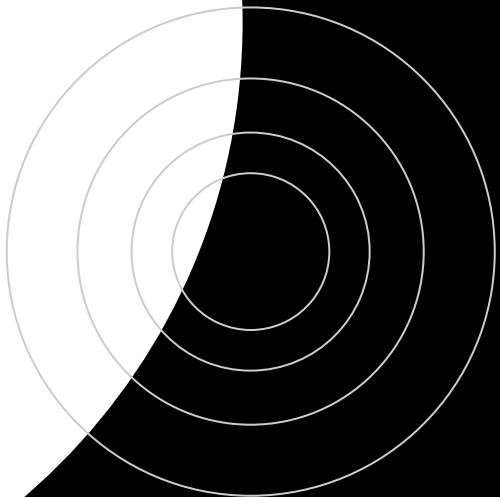


1

Качество

модели ALS

F1-score = 0.000040



Базовая модель ALS

"Implicit Feedback"! – отсутствие рейтингов

ALPHA = 40

Features
=200

F1-Score = 0.00558





2

Скорость

подбора рекомендаций

Collaborative Filtering (CF)

BPR = Bayesian Personalized Ranking mode

```
model_bpr = BayesianPersonalizedRanking(  
    factors=200, learning_rate=0.3,  
    iterations=500,  
    regularization=0.01)
```

Все книги:

Hit-Rate_avg: 0.0207
Recall_avg: 0.00114
Precision_avg: 0.00719
F1-score_avg: 0.00197

Книги,
у которых > 1 читателя:

Hit-Rate_avg: 0.1102
Recall_avg: 0.00667
Precision_avg: 0.06128
F1-score_avg: 0.01203

*Factorization
machines*

Loss = BPR



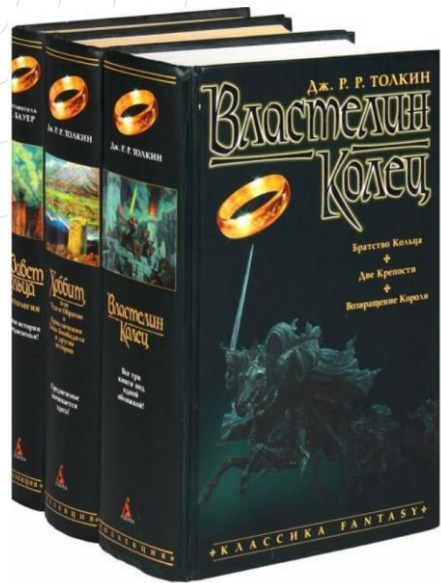
#RankFM
#Implicit



Неявные дубликаты



Похожие книги



Разные годы
издания



Книги из
одного сборника

! Точность алгоритма ALS



4

Парадокс

Предсказания VS Рекомендации

Новые авторы

100% Hit Rate – угадали все произведения Анны Тодд

Новизна рекомендаций ☹



! Предсказания новых авторов с помощью ALS-модели





3

Неактивные читатели

Content Based Filtering (CB)



- Жанры
- Авторы
- Названия

“Item-Item”
recommendations

Cosine
similarity



#TF-IDF
#nltk

Книги похожих жанров

```
print("Сложные bbk коды, состоящие из нескольких блоков:")  
df_ML[df_ML['bbk_ml'].str.contains(' ')].bbk_ml.head(20)
```

Сложные bbk коды, состоящие из нескольких блоков:

```
0          ШЗ(257=Ки)-611,0 Ш163.25-314,0  
4          ТЗ(5Ид)63-283,02 ТЗ(5Ид)63-5,02  
6  Ш5(2=Р)43-318.3,0 Ш5(2=Р)51-318.3,0 Ш5(0)32-31...  
8          Ш401.171.11,0 Ш5(5Вл)42-35,0
```

Ш Филологические науки. Художественная литература

> Ш,0 Общие работы по филологии

> Ш.а/я Общий раздел

> Ш0 Частные филологии. Текстология

> Ш1 Языкознание

> Ш2/3 Фольклор (народнопоэтическое творчество). Фольклористика

<https://lod.rsl.ru/bbkqsk/concepts/%D0%AE9>





5

Ресурсы

для similarity matrix

```

def fast_cosine_similarities(sparse_matrix, axis_type=1):

    # матрица mxn
    # m - объекты, n - признаки
    # нормализуем все строки или столбцы, в зависимости от axis
    csc_matrix = pp.normalize(sparse_matrix.tocsc(), axis=axis_type)

    transposed = csc_matrix.T
    # расстояния между объектами-строками
    # матрица mxm
    if axis_type == 1:
        return csc_matrix * transposed
    # расстояния между признаками-столбцами
    # матрица nxn
    else:
        return transposed * csc_matrix

# sim_matrix = fast_cosine_similarities(tfidf, axis_type=1)
# sim_matrix

```

```

<119805x119805 sparse matrix of type '<class 'numpy.float64''>'
  with 676185021 stored elements in Compressed Sparse Row format>

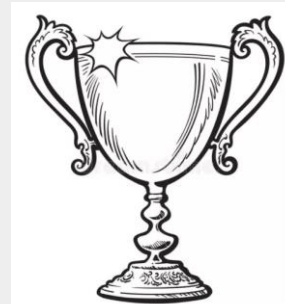
```

Гибридные рекомендации

ALS (CF)
активные
читатели

TF-IDF(CB)
неактивные
пользователи

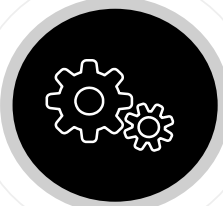
F1-Score = 0.0072



(CB) – похожие книги (жанр, название, автор)

(CF) – книги читателей с похожим профилем

Результат проекта



Гибридное
решение



Collaborative
Filtering



Content-Based
Filtering



Спасибо!

Мои контакты:

Дарья Хафизова,

+79172495796,

sunsaffron@yandex.ru