Recommender System

User’s review Amazon products

https://github.com/antoniosh97/Recommender-System-2023

Recommender System - Our Team

“A” Team

Antonio

Brenda

Evaristo

Joan

Recommender System – 1. Motivation

The ultimate goal of a recommender systems is to **unlock the maximum potential of customer purchases,** making them a crucial component of a successful personalized marketing strategy.

Improve user experience: help users to discover new products and services based on their own interests.

User **loyalty**: encourage customers to return.

User engagement: encourage users to **engage with the content** based on the recommendation of the products or services.

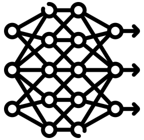
More direct and **personalized marketing campaigns**

Recommender System - 2. Goals

Baseline

Propose **two simple approaches** to give a recommendation: **most popular products (1) and random products (2).**

Define and develop the initial ETL and data preparation for modelling.



Deep learning models

Develop two different **deep learning approaches for the recommender system: Factorization machines (FM) and Neural matrix Factorization model (NeuMF).**

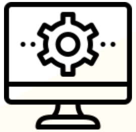
Perform ablation experiments to give an intuition on hyperparameter tuning.

Business sense of a recommender system

Understand the **performance** of the different models based on a **business perspective.**

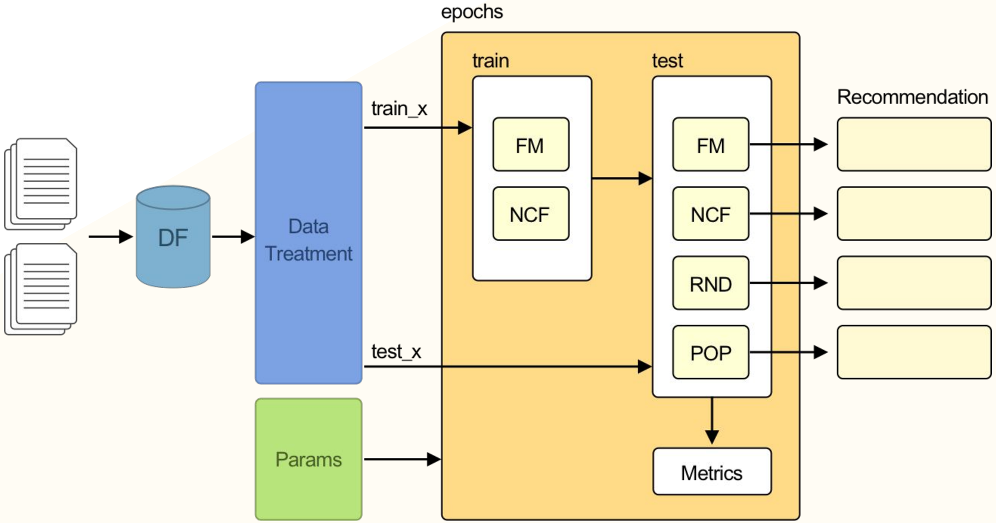
Conclude the project with **next steps** to perform in the future.

Recommender System - 3. Proposal



Strategy

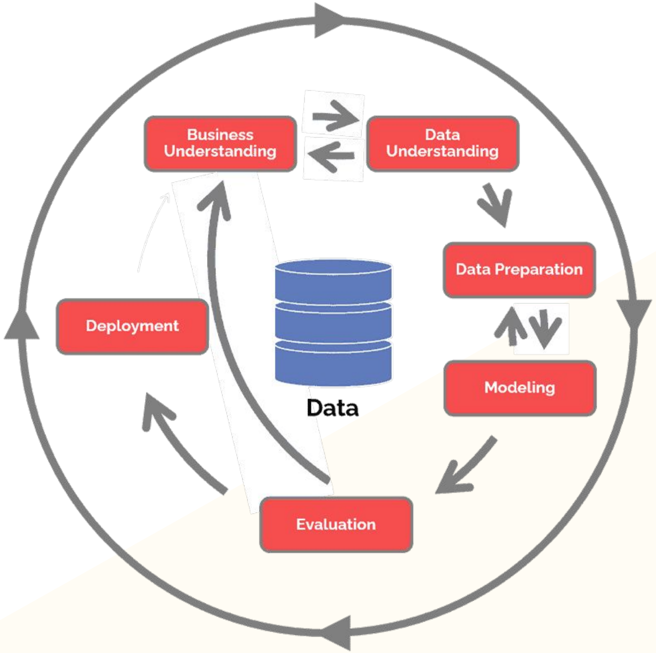
1. Models: FM, RND, POP, NCF
2. Different Datasets: Amazon and MovieLens
3. Data transformation and Sampling strategy
4. Training and Test
5. Metrics: HR, NDGC, Coverage
6. Histograms and Graphics: Plot and Tensorboard
7. Tuning parameters
8. Custom parametrization



Pipeline

Recommender System - 3. Proposal

Methodology



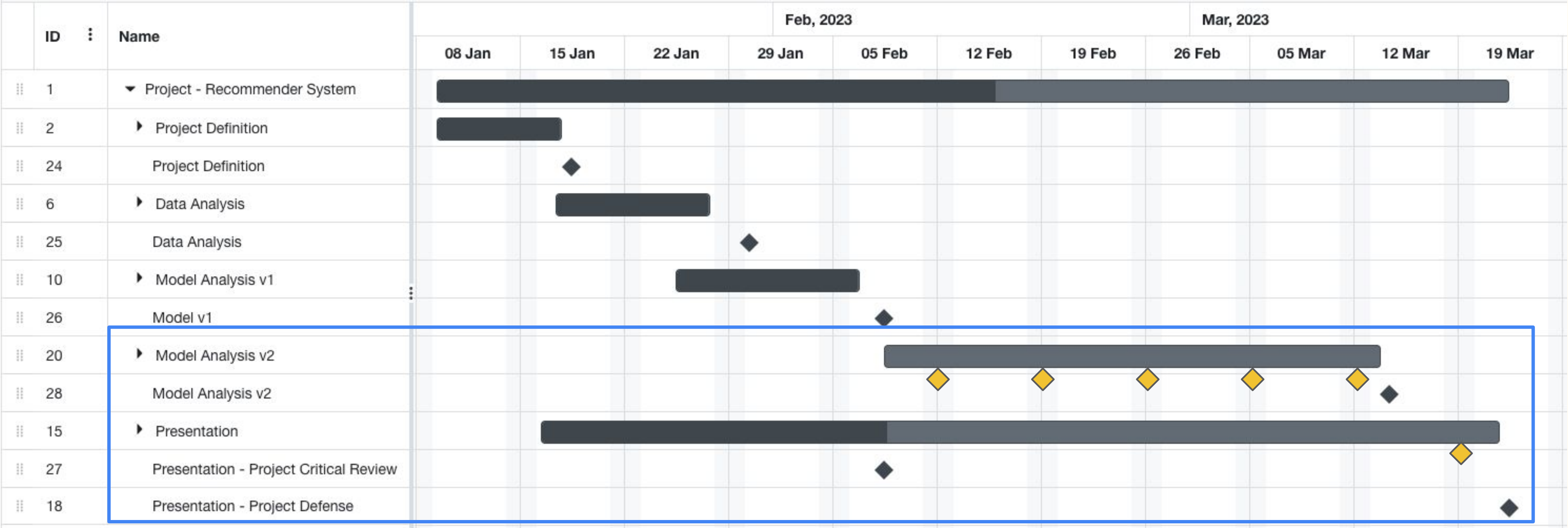
CRISP-DM
Cross Industry Standard Process for Data Mining

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

Waterfall

Recommender System - 3. Proposal

Project Plan and Milestones



Recommender System - 4. Implementation

Datasets

Amazon: Original

Total Reviews
233.1 million

Subset: Musical Instruments

Reviews	Products	Users
1,512,530	112,222	903,330

MovieLens-100k

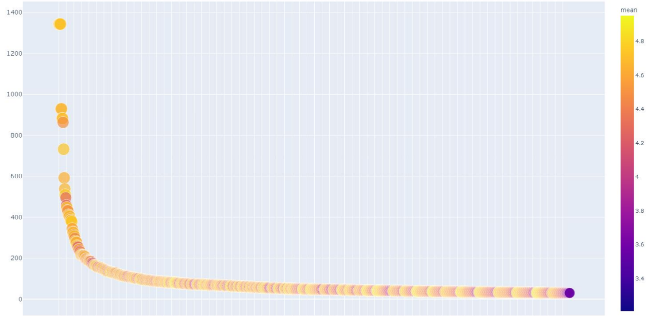
Reviews	Movies	Users
100.000	1,682	943

Structure: User, Item, Review(1-5), Timestamp

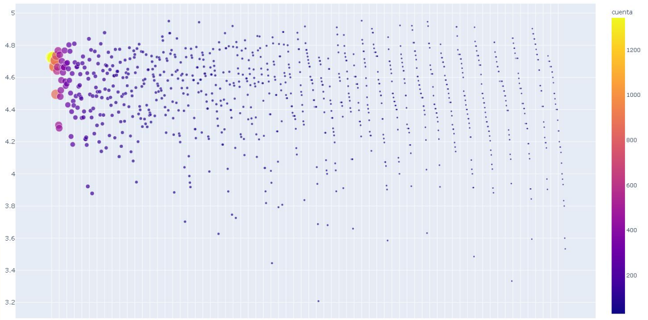
Recommender System - 4. Implementation

Datasets

Amazon

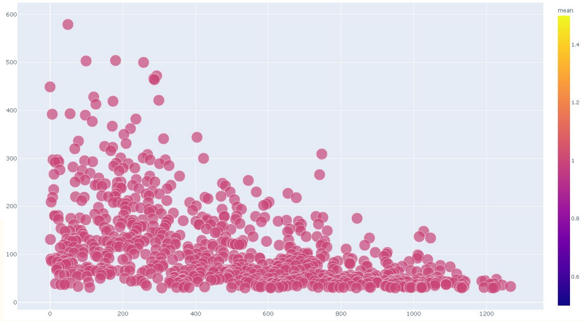


1 . Reviews for Items: >29, 952 of 137364

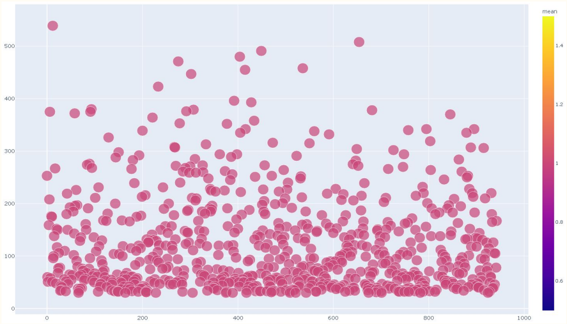


2 . Reviews 1-5: >29, 952 of 137364

MovieLens



3 . Reviews for Items: >29, 806 of 94443



4 . Reviews by User: >29, 732 of 94443

Recommender System - 4. Implementation

Datasets

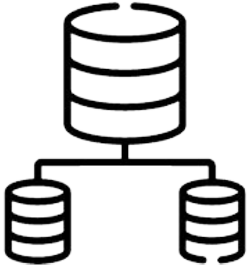
Data: Transformation and cleaning

- Format data types
- Remove duplicates
- Reduce reviews by user and item

	Amazon		**MovieLens**
Users	4.275	14.138	943
Items	2.080	6.178	1.682
Reviews	52.023	137.364	100.000
RxUxI	8x8	6x6	20 reviews

Recommender System - 4. Implementation

Sampling



Training

Implicit feedback (1-5 -> 1)

Negative Sampling (k=4-6)

Test

TLOO Time Leave One Out (x=1)
RLOO Random Leave One Out (x=1)

Full Ranking
(add neg all items/user)

Recommender System - 4. Implementation

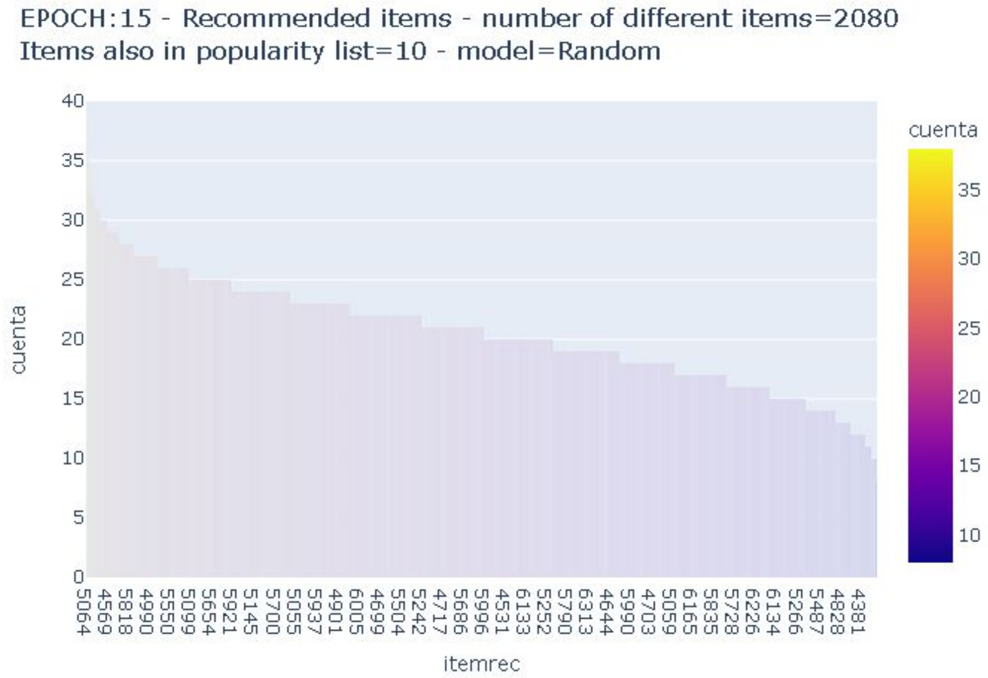
Models

- Random
- Popularity
- FM - Factorization Machine
- NCF - Neural Collaborative Filtering

Recommender System - 4. Implementation

Models

Random

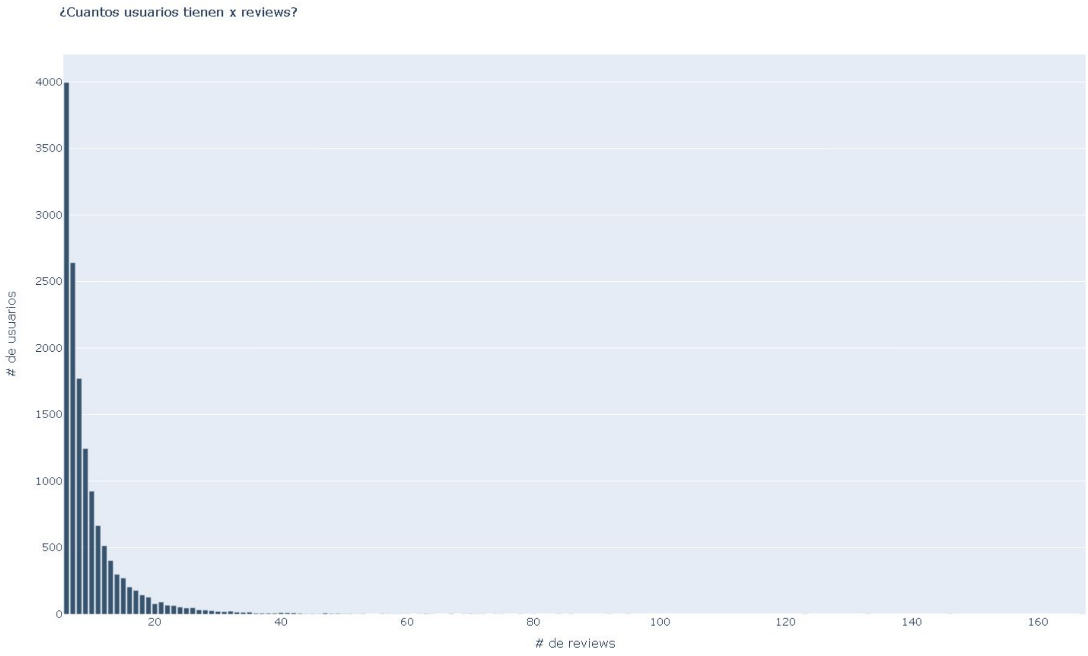


- Recommend items of the top@k list randomly
- Has a coverage of 100% of the products set

Recommender System - 4. Implementation

Models

Popularity

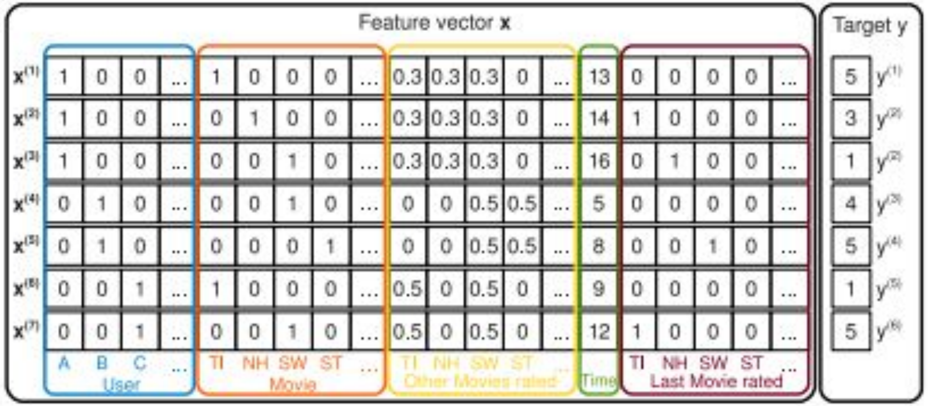


- Most simple statistic method
- Select the first top@K items with more ratings
- Solves the cold start problem with new users but always recommends the same set of products

Recommender System - 4. Implementation

Models

FM - Factorization Machine

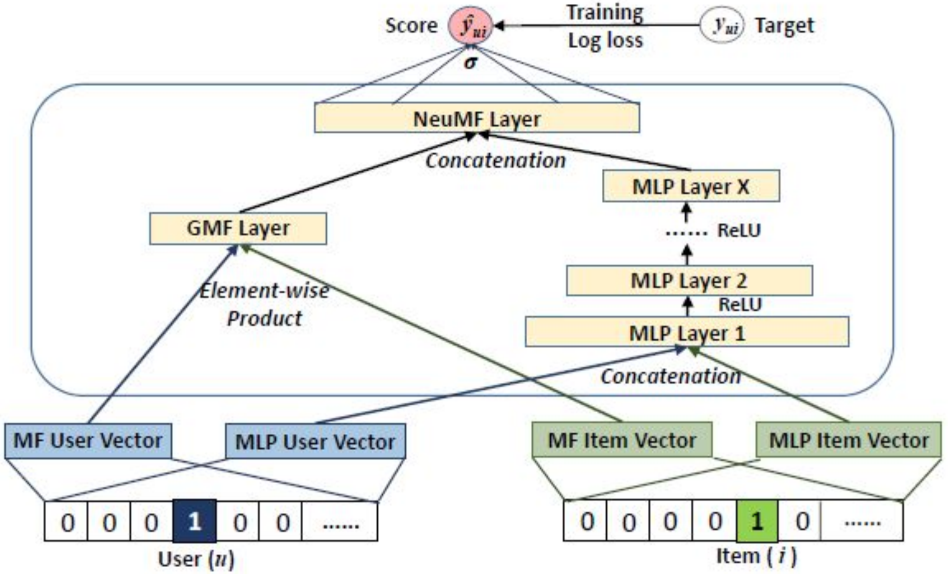


Implement Factorization Machine to perform a ranking prediction task of top@k

Recommender System - 4. Implementation

Models

NeuMF



- We implemented NeuMF Neural Matrix Factorization model that combines MF with Neural MLP
- Changed the original last MLP layer activation function SIGMOID followed by BCELoss to use only a BCEwithLogitsLoss function in train

Recommender System - 4. Implementation

Metrics

Hit Ratio HR@topk.
How often is the GT item present in the recommendation list

NDCG Normalized Discounted Cumulative Gain.
Measures ranking quality

Coverage compares the set of total recommended items with the total items in train dataset

Recommender System - 4. Implementation

Computational resources

- GIT repository (recommender system project AIDL2023)
- Colab notebooks
- CPU (local)
- Tensorboard
- Miniconda environment
- Process average time: 2h

Recommender System – 5. Results. Ablation studies

Illustrative for NeuMF and TLOO

	K negative sampling	***Learning rate***	***Hidden size***	***Metrics***
Exp 1	*5*	*1 e-3*	*32*	*HR: 0.05 | NDCG:0.02 | COVERAGE: 68.2*
Exp 2	*6*	*1 e-4*	*64*	*HR: 0.07 | NDCG:0.04 | COVERAGE: 16.2*
Exp 3	*6*	*1 e-4*	*32*	*HR: 0.06| NDCG:0.03 | COVERAGE: 16.3*
…	**…**	**…**	**…**	**…**

To determine the best hyperparameters, a grid is created with different configurations. The performance is measured in every iteration and the final model is chosen from the best of these experiments.

Recommender System – 5. Results. Final configuration

	K negative sampling	Learning rate	Hidden size	Metrics
Best exp NeuMF	*5*	*1 e-4*	*64*	*HR: 0.07 | NDCG:0.04 | COVERAGE: 35.8*
Best exp FM	*5*	*1 e-3*	*64*	*HR: 0.06 | NDCG:0.03 | COVERAGE: 49.6*
Random	-	-	-	*HR: 0.004 | NDCG:0.002 | COVERAGE: 100*

NeuFM model has the best metrics in terms of HR and NDCG. With the deep learning approach we have an x10 improvement from the random recommendation.

Recommender System – 6. Conclusions

Hypothesis accomplished
Key points in REC
Compliance in business

Project Conclusions

Add context
Bias in dataset and models
Control and insertion of bias
Docker & Flask

Next steps

Transparency
Limits
Privacy

Ethical

Recommender System

Thank you! Questions?

https://github.com/antoniosh97/Recommender-System-2023