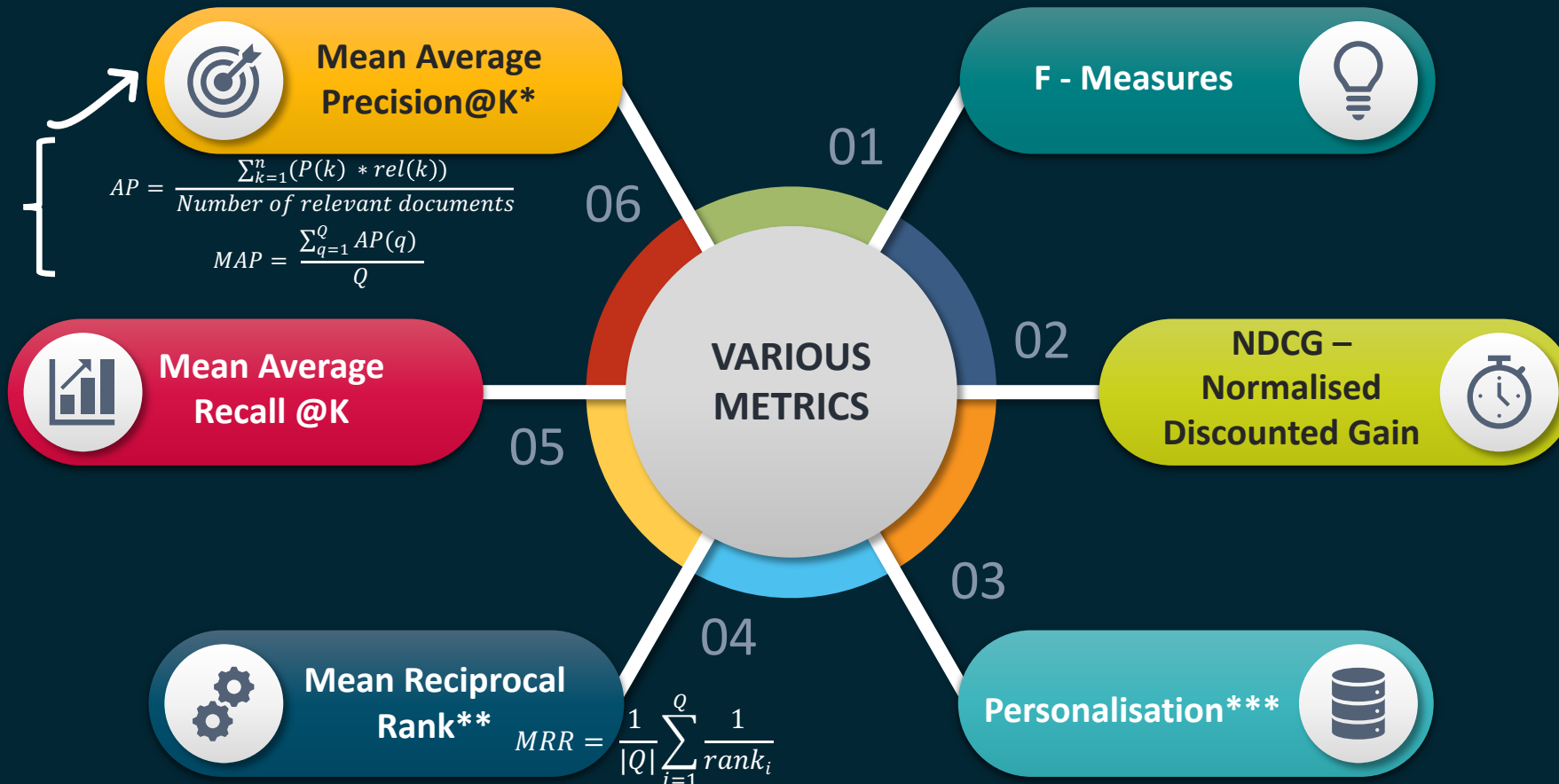


Evaluation Metrics for Recommender System

Overview

Recommender systems are probably the most popular forms of application of machine learning/deep learning or Reinforcement Learning algorithms which has perhaps intrigued both the product owners and researcher community in terms of the critical challenges it has posed to them. For RS, many algorithms have been suggested for providing recommendations and this has led to considerable amount of debates among DS community in terms of the choice of the most appropriate evaluation techniques to measure the efficacy of various algorithms. Here , we present a list of some of the most popular methods, that are being used to evaluate the relative strength of such algorithms.



Other Metrics

- 🔍 Coverage
- 🔍 Intra-list Similarity
- 🔍 Novelty
- 🔍 Recommender Precision
- 🔍 Recommender Recall
- 🔍 Root Mean Squared Error
- 🔍 Mean Absolute Error
- 🔍 Spearman's Rho
- 🔍 Kendell's Tau

* For MAP@K, the 'Q' represents the number of queries, AP is the average precision , P(k) is the precision at top-k recos and rel(k) is an indicator equaling 1 if the item at rank is a relevant document, and zero otherwise.

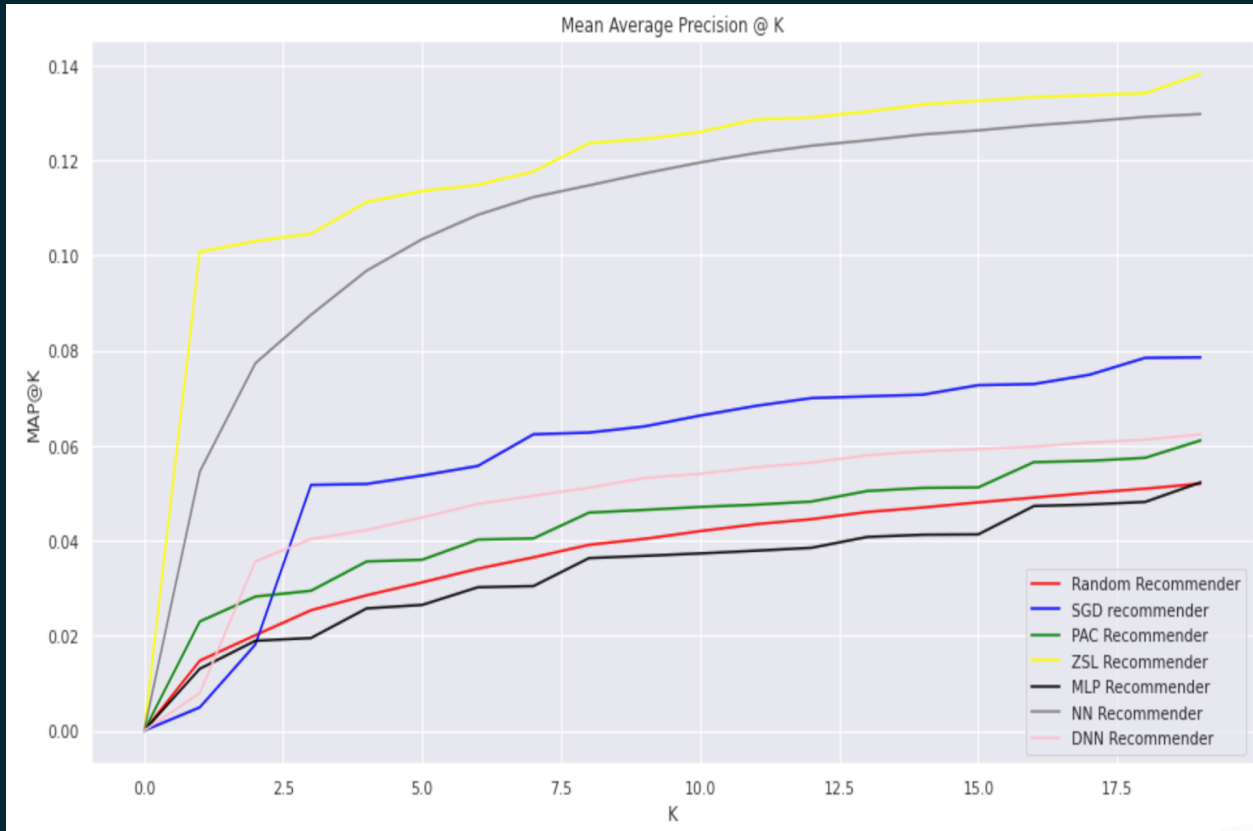
**For MRR, the 'Q' represents the number of queries, and it is an aggregate over all the queries made.

*** For our use case (dealing with the Cold Start problems)– We will only report - MAP@K, MAR@K , Personalization Index value, Overall Recommender Recall, Overall Recommender Precision for now.

Evaluation Summary Stats: ZSL and other algorithms* - User Cold Start



MAP@K



ZSL is the clear winner while measuring the Mean Avg Precision @K measured for all the 'pseudo cold' customers, while recommending warm items, in comparison with the available ground truth based on past responses of the customers



Personalization Score*

Algorithms	Personalization Score
ZSL technique	0.45
Multi-layer Perceptron	0.0004
Passive Aggressive Classifier	0.019
KNN Based Approach	0.62
Random Allocation	0.35
SGD Classifier	-2.220446049250313e-16



Reco. Precision/Recall Score**

Algorithms	Precision	Recall
ZSL technique	0.06	0.80
Multi-layer Perceptron	0.022	0.38
Passive Aggressive Classifier	0.021	0.39
KNN Based Approach	0.05	0.93
Random Allocation	0.01	0.58
SGD Classifier	0.023	0.40

*/**Personalization Score refers to the degree of dissimilarity(1- cosine similarity) between cold users' list of recommendation, MAP@K reveals how relevant the list of recommended items are and we also have reported the recommender level overall precision and recall numbers as well.