Mean Reciprocal Rank (MRR)

🔸 Definition

Mean Reciprocal Rank (MRR) is a statistical evaluation metric used to assess the performance of systems that return a ranked list of items, such as search engines, question-answering systems, and recommender systems. It measures how highly the first correct or relevant result appears in the output list for each query. The “reciprocal rank” for a single query is the inverse (1 divided by) of the rank at which the first relevant result appears. MRR then computes the average of these reciprocal ranks across all queries.

Key Insight: Higher MRR values (closer to 1) indicate that relevant results are appearing earlier in the ranking, which is desirable.

🔸 Formula

MRR = (1 / |Q|) * Σ (1 / rank_i), for i = 1 to |Q|

Where:
- Q: Total number of queries
- rank_i: The rank position of the first correct (i.e., relevant) result for the i-th query
- 1 / rank_i: The reciprocal rank for query i

Note: If no relevant result is found for a query, its reciprocal rank is considered as 0.

🔹 Step-by-Step Example

Let’s evaluate a system's performance for 3 queries, each returning a list of documents:

Let D4 be a relevant doc.

Query	Returned List	First Relevant Document Rank
Q1	[D1, D4, D2]	2
Q2	[D4, D2, D1]	1
Q3	[D5, D3, D1]	— (no correct document)

Reciprocal Ranks:
- Q1 ⟶ 1/2 = 0.5
- Q2 ⟶ 1/1 = 1.0
- Q3 ⟶ 0 (no relevant document found)

MRR = (1/3) * (0.5 + 1.0 + 0) = 1.5 / 3 = 0.5

🔹 Applications of MRR

MRR is widely applicable in fields where ranked retrieval is important and the first correct response holds primary importance:

1. Search Engines: Ranking documents/webpages in response to a search query.
2. Question Answering (QA) Systems: Retrieving and ranking possible answers from a corpus.
3. Dialogue Systems / Chatbots: Evaluating how early in the conversation list the correct intent or response appears.
4. Recommender Systems: Especially useful when only the top recommendation matters.
5. Information Retrieval (IR) Benchmarking: Used in TREC, MS MARCO, BioASQ, etc.

🔹 Characteristics

- Range: 0 to 1
  - 1: Perfect — correct answer always ranked first.
  - 0: Worst — no correct answers ever found.
- Simple and Interpretable: Easy to compute and explain.
- Sensitive to Top Results: Prioritizes early retrieval of relevant results.