**Title: Evaluating a Movie Recommendation System Using Precision, Recall, and F1-Score (Real-Time Scenario)**

**🔷 1. Introduction**

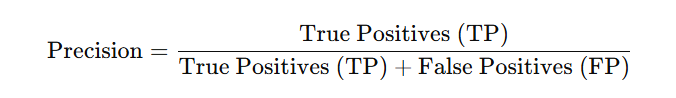
In the age of digital entertainment, platforms like **Netflix**, **Amazon Prime**, and **Hotstar** use **recommender systems** to suggest content tailored to each user. However, to know how **effective and relevant** these recommendations are, we must evaluate the system's performance using **online evaluation metrics**.

Unlike **offline evaluation** (which uses historical data), **online evaluation** involves **real-time data** collected from **user behavior** — such as watching, clicking, or rating a movie.

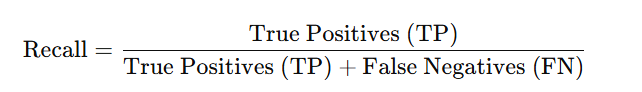
**🔷 2. Key Metrics in Online Evaluation**

To assess how accurate the recommendations are, we use the following three main metrics:

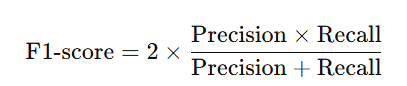
**✅ Precision**

* Measures the **accuracy** of recommendations.
* Formula:
* 
* **High precision** means most recommended items were relevant.

**✅ Recall**

* Measures the **coverage** — how many relevant items were recommended.
* Formula:
* **High recall** means few relevant items were missed.

**✅ F1-Score**

* Combines precision and recall into a single metric.
* Formula:
* **High F1-score** means balanced performance.

**3. Real-World Scenario – Netflix Example**

Let’s take a real-time example of a user named **Ravi** using **Netflix**.

**🔸 Ravi logs into Netflix on Sunday evening.**

Netflix recommends these 5 movies to him:

**Recommended Movies:**  
[The Matrix, Dangal, Interstellar, Hera Pheri, Titanic]

Ravi watches or rates these 3 movies during that session:

**User's Actual Interactions:**  
[The Matrix, Interstellar, Inception]

**🔷 4. Step-by-Step Evaluation**

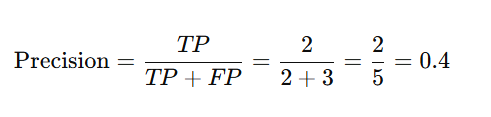
Let’s now evaluate how effective Netflix's recommendations were for Ravi.

**🔸 Step 1: Compare the Lists**

* **True Positives (TP)**: Recommended AND watched  
  → The Matrix, Interstellar → ✅ **2**
* **False Positives (FP)**: Recommended but NOT watched  
  → Dangal, Hera Pheri, Titanic → ❌ **3**
* **False Negatives (FN)**: Watched but NOT recommended  
  → Inception → ❌ **1**

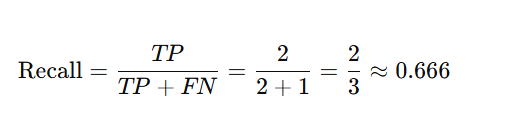
**🔸 Step 2: Apply the Formulas**

**🔹 Precision:**



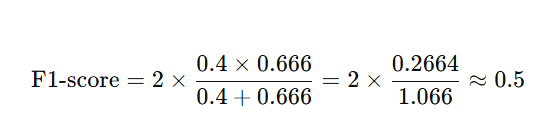
📝 This means: Out of all 5 recommended movies, only 2 were actually relevant.

**🔹 Recall:**



📝 This means: Out of the 3 movies Ravi actually liked, Netflix managed to recommend 2.

**🔹 F1-Score:**



📝 This shows a balanced but moderate recommendation performance.

**🔷 5. Interpretation and Analysis**

**🔸 Precision Analysis**

Netflix gave 5 movie recommendations, but Ravi found only 2 relevant — indicating **moderate precision (40%)**.

**🔸 Recall Analysis**

Ravi interacted with 3 movies total, and Netflix captured 2 — so the **recall is high (66%)**, meaning most relevant movies were suggested.

**🔸 F1-Score Analysis**

Combining both metrics gives us an F1-score of **0.5**. This is a balanced score, showing that while some irrelevant movies were recommended, the system still performed fairly well.

**🔷 6. Summary Table**

| **Category** | **Movies** | **Count** |
| --- | --- | --- |
| Recommended Movies | The Matrix, Dangal, Interstellar, Hera Pheri, Titanic | 5 |
| Watched Movies | The Matrix, Interstellar, Inception | 3 |
| True Positives (TP) | The Matrix, Interstellar | 2 |
| False Positives (FP) | Dangal, Hera Pheri, Titanic | 3 |
| False Negatives (FN) | Inception | 1 |

| **Metric** | **Formula** | **Value** |
| --- | --- | --- |
| Precision | TP / (TP + FP) = 2 / 5 | 0.4 |
| Recall | TP / (TP + FN) = 2 / 3 | 0.666 |
| F1-Score | 2 × (P × R) / (P + R) ≈ 2 × 0.266 / 1.06 | ≈ 0.5 |

**🔷 7. Conclusion**

In this **real-time evaluation** of a Netflix movie recommendation system:

* **Precision was 40%**, which means some recommendations were not helpful.
* **Recall was 66.6%**, meaning the system did well in suggesting movies Ravi liked.
* **F1-score of 0.5** shows a **moderate but balanced** performance.

This type of evaluation helps platforms like Netflix to **continuously improve** their recommender systems by learning from **user interactions** in real-time.