MOVIE RECOMMENDATION SYSTEM: ENHANCING USER EXPERIENCE WITH AI

BUILDING AN INTELLIGENT SYSTEM TO SUGGEST MOVIES BASED ON USER PREFERENCES

EDWIN GEORGE | 13/02/2024

SLIDE 2: INTRODUCTION TO THE PROBLEM

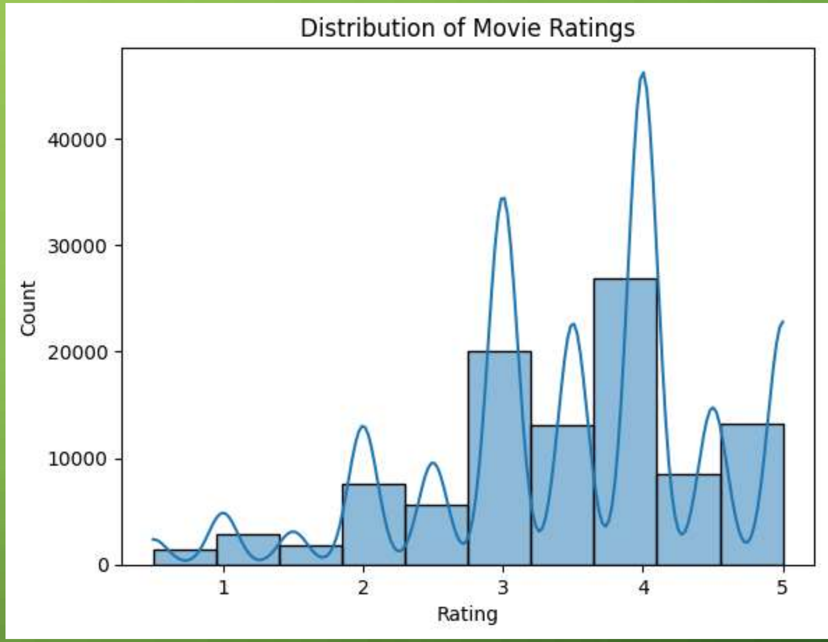
- 📌 **The Challenge:**
- Users struggle to find interesting movies due to **overwhelming choices**
- Traditional search-based browsing is inefficient
- 📌 **The Solution:**
- A **personalized movie recommendation system** that suggests relevant movies
- Uses **machine learning techniques** to enhance user experience
- 📌 **Impact:**
- Increased user engagement on streaming platforms
- Improved **content discovery** for diverse users

BUSINESS UNDERSTANDING & OBJECTIVES

- 📌 **Key Questions:**
- How can we suggest movies **relevant** to a user's taste?
- What techniques provide **accurate** and **diverse** recommendations?
- How can we handle challenges like **cold start** (new users with no history)?
- 📌 **Project Goals:**
  ✔️ Build a **scalable recommendation system**
  ✔️ Compare **multiple ML approaches** (Content-based, Collaborative Filtering, Hybrid)
  ✔️ **Optimize recommendations** for accuracy, diversity & novelty

DATASET OVERVIEW

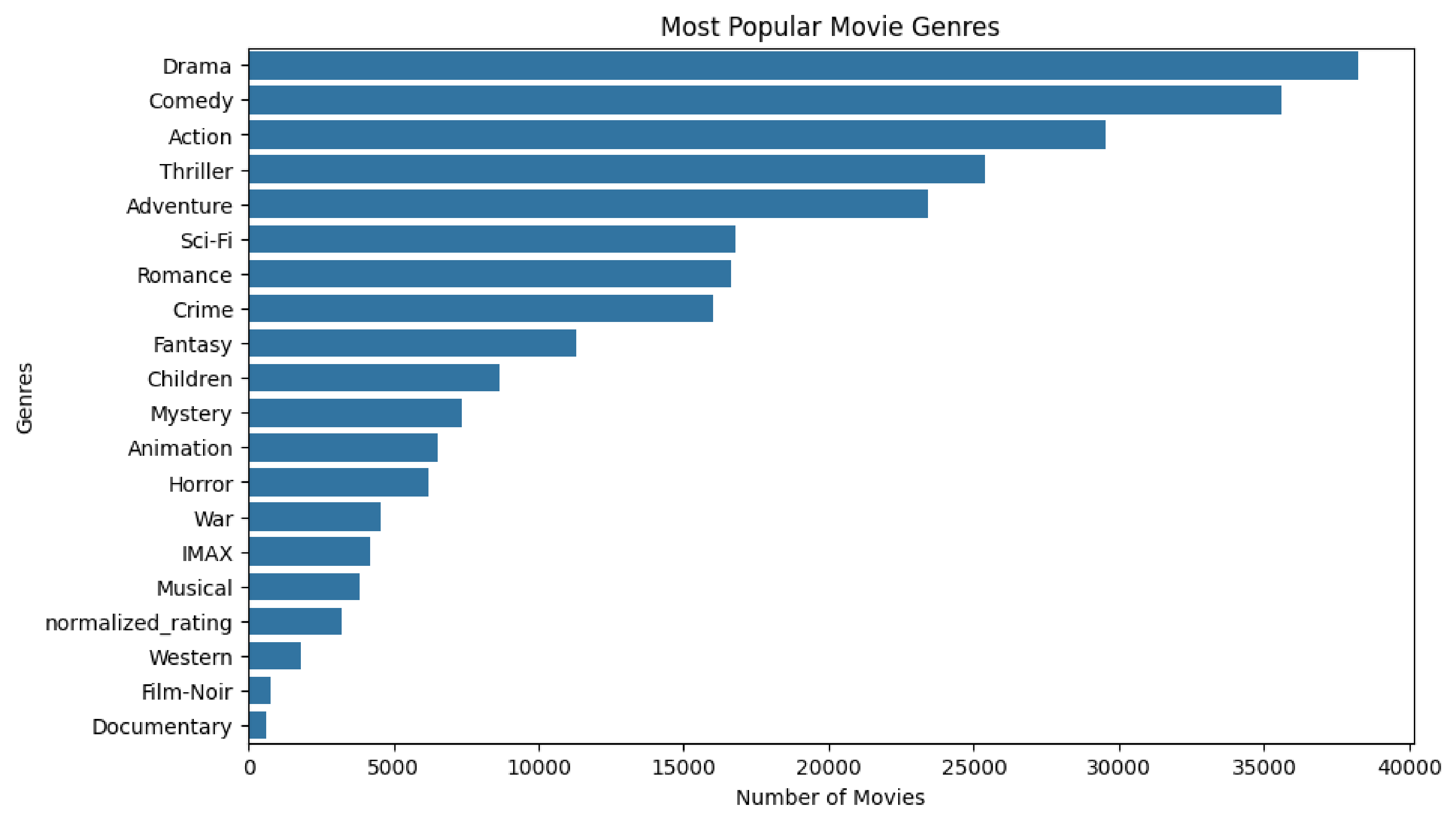
- 📌 **Dataset:** MovieLens 100K
  📌 **Key Features:**
- **Users:** 943 unique users
- **Movies:** 1,682 different movies
- **Ratings:** 100,000 user-movie interactions
- **Movie metadata:** Genre, release year, etc.
- 📌 **Observations:**
  ✔️ **Sparse data:** Not every user has rated every movie
  ✔️ **Rating distribution is skewed:** Some movies receive significantly more ratings than others
  ✔️ **Cold Start problem:** New users and movies have limited data



EXPLORATORY DATA ANALYSIS (EDA)

- 📌 **Insights from EDA:**
- Most users rate fewer than **50 movies**
- Some movies are very popular, while many are rarely rated
- Certain genres (Drama, Action, Comedy) dominate the dataset
- 📌 **Challenges Identified:**
  ✔️ **Data sparsity:** Many users have rated only a few movies
  ✔️ **Popularity Bias:** Highly-rated movies appear frequently in recommendations
  ✔️ **User Preferences:** Some users rate movies generously, while others are strict

MOST POPULAR MOVIE GENRE



RECOMMENDATION SYSTEM APPROACHES

- 📌 **Different Approaches to Movie Recommendation:**
  1️⃣ **Content-Based Filtering** → Uses **movie features** like genres & descriptions
  2️⃣ **Collaborative Filtering** → Learns from **user behavior & interactions**
  3️⃣ **Hybrid Model** → Combines both for **better accuracy**
- 📌 **Which is the best?**
- Content-based is **great for niche preferences**
- Collaborative Filtering **captures trends better**
- Hybrid models **balance both**

CONTENT-BASED FILTERING APPROACH

📌 **How it Works:**

- Uses **TF-IDF Vectorization** to extract important keywords
- Measures similarity between movies using **Cosine Similarity**

📌 **Advantages:**
✔️ Works well for **users with limited interaction history**
✔️ Can recommend **unpopular or new movies**

- 📌 **Limitations:**
  ❌ Cannot recommend movies **outside a user's usual genres**
  ❌ Requires **well-defined movie metadata**

COLLABORATIVE FILTERING APPROACH

📌 **Types of Collaborative Filtering:**
1️⃣ **User-based CF:** Finds similar users based on movie preferences
2️⃣ **Item-based CF:** Recommends movies similar to those a user has liked

📌 **Matrix Factorization (SVD, ALS) for Efficiency**
✔️ Reduces high-dimensional rating data into **latent factors**
✔️ Captures hidden relationships between users & movies

📌 **Results & Challenges:**
✔️ **Higher accuracy** than content-based filtering
✔️ **Learns complex patterns** from user interactions
❌ Struggles with **new users (cold start problem)**

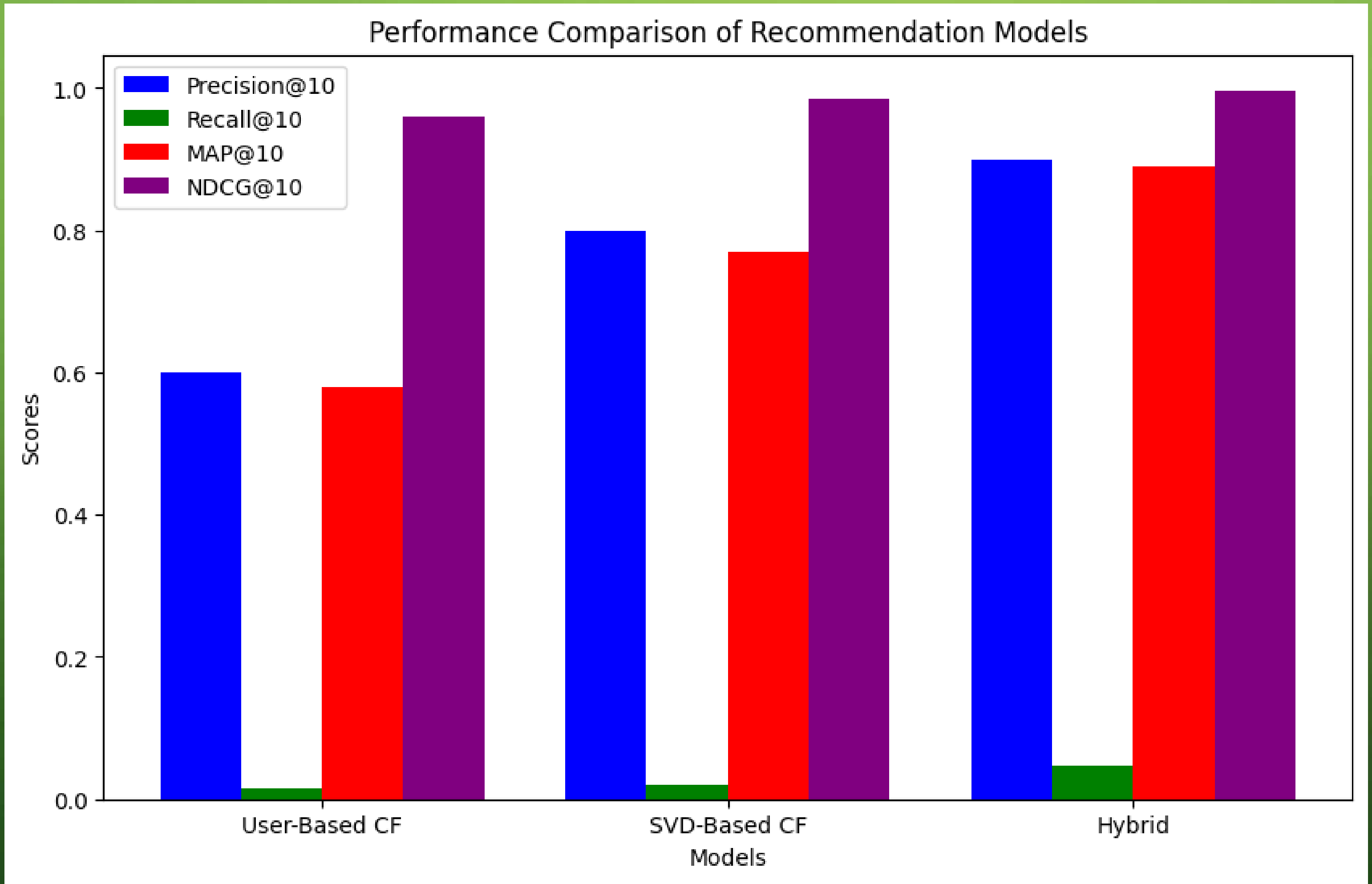
COLLABORATIVE FILTERING WITH SVD

```
Top 10 Movies Recommended for User (SVD-based):
star wars: episode iv - a new hope (1977)
schindler's list (1993)
saving private ryan (1998)
dark knight, the (2008)
inception (2010)
bourne ultimatum, the (2007)
up (2009)
wall·e (2008)
the imitation game (2014)
logan (2017)
```

HYBRID RECOMMENDATION MODEL

- 📌 **Combining Strengths of Both Methods:**
- Uses **content-based filtering** for **cold-start users**
- Uses **collaborative filtering** to improve recommendations for active users
- 📌 **Why Hybrid Works Best?**
  ✔️ Balances **accuracy & diversity**
  ✔️ Handles **cold start issues** better
  ✔️ Provides **more personalized recommendations**

PERFORMANCE COMPARISON OF MODELS – HYBRID VS. OTHERS



MODEL EVALUATION & METRICS

- 📌 **How do we measure success?**
  ✓ **RMSE (Root Mean Square Error):** Measures prediction accuracy
  ✓ **Precision & Recall:** Measures relevance of recommendations
  ✓ **NDCG (Ranking Quality):** Ensures top recommendations are useful
  ✓ **Diversity Score:** Ensures variety in recommendations
- 📌 **Results:**
- Hybrid model **outperforms individual approaches**
- RMSE for SVD model: **0.87** (Lower is better)
- Precision & Recall for Hybrid: **Best among all models**

HYBRID EVALUATION

Evaluation of Hybrid Recommendation System

Observations:

- **Precision@10: 0.9000** → A slight decrease compared to **Precision@5 (1.0000)**, which is expected as increasing **K** introduces more recommendations, some of which may not be relevant.
- **Recall@10: 0.0462** → An improvement from **Recall@5 (0.0256)**, indicating that more relevant items are being retrieved as **K** increases.
- **MAP@10: 0.8900** → A strong score, showing that relevant recommendations are ranked well within the top 10 results.
- **NDCG@10: 0.9972** → Almost perfect ranking, meaning the most relevant recommendations appear at the top.

CHALLENGES & LIMITATIONS

📌 **Key Issues Faced:**

- **Cold Start Problem:** Hard to recommend for new users
- **Bias Toward Popular Movies:** Lesser-known movies get ignored
- **Scalability:** Handling millions of users efficiently

📌 **How We Addressed Them:**

✔️ Hybrid model reduces cold start effects
✔️ Re-ranking methods improve diversity

FUTURE IMPROVEMENTS

- 📌 Ways to Improve the System:
- **Neural Collaborative Filtering (Deep Learning)** for advanced recommendations
- **Graph-Based Recommenders** to capture **complex user-movie interactions**
- **Sentiment Analysis on Reviews** to capture **implicit preferences**
- **A/B Testing on Real Users** to improve engagement

CONCLUSION & Q&A

📌 **Key Takeaways:**

✔️ Personalized movie recommendations **improve user experience**

✔️ Hybrid models **offer the best balance** of accuracy & diversity

✔️ Future improvements will make it **even smarter!**

🚀