# Evaluating Recommender Systems FEVR: Implementation and Analysis using the Netflix Dataset

Musaed owied – 2140988, Elyas Sindi – 2143512, Nawaf fayez alrehaili -2142609

## Abstract

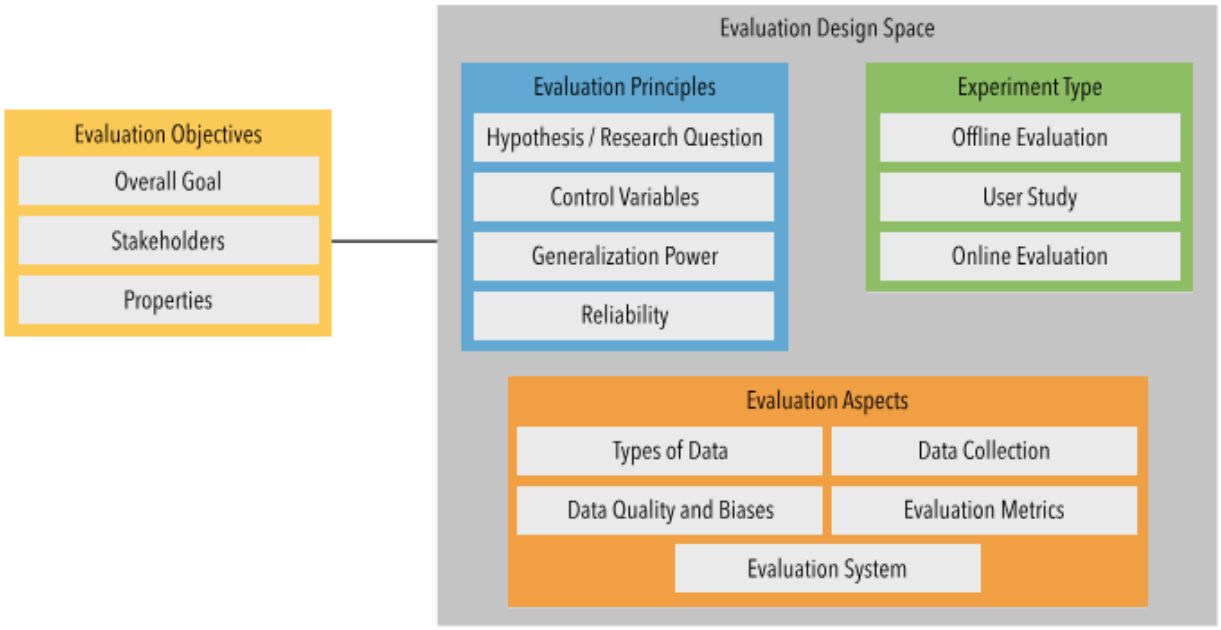
The comprehensive evaluation of recommender systems is crucial for their effectiveness in various applications. This report details the implementation and evaluation of recommendation algorithms using the Netflix dataset, guided by the FEVR framework introduced by Eva Zangerle and Christine Bauer. Our approach involves testing SVD-based models and item-based Pearson correlation methods, focusing on prediction accuracy and user satisfaction. Key metrics such as RMSE and MAE are used to evaluate performance. The results demonstrate the strengths and limitations of each method, providing insights for future improvements in recommender system evaluation.

## Introduction

Recommender systems are integral to personalized content delivery, influencing user satisfaction and engagement. This project aims to implement and evaluate different recommendation algorithms using the Netflix dataset. We focus on SVD-based models and item-based Pearson correlation methods, analyzing their performance through key metrics such as RMSE and MAE. Our goal is to validate the FEVR framework's applicability in comprehensive recommender system evaluation and provide insights into optimizing recommendation accuracy.

## Background/Related Work

The FEVR framework, introduced by Zangerle and Bauer (2023), emphasizes a multi-faceted approach to evaluating recommender systems. This survey paper provides a comprehensive overview of evaluation methods, summarizing and synthesizing existing literature to give readers a broad understanding of the state-of-the-art, key developments, trends, and gaps in the research area. The FEVR framework categorizes essential facets such as evaluation objectives, principles, and metrics. It addresses the rich evaluation design space, including evaluation setup, data collection, and employed metrics. The primary objective of this framework is to consolidate and systematically organize dispersed knowledge on recommender system evaluation, providing a structured foundation for comprehensive evaluations.




FEVR framework

## Approach

Our approach involves implementing SVD-based and item-based Pearson correlation recommendation algorithms. We selected these methods due to their proven effectiveness in previous research. The SVD model aims to capture latent factors in user-item interactions, while the item-based Pearson correlation method computes similarities between items. We used RMSE and MAE as evaluation metrics to measure predictive accuracy. The Netflix dataset, with its extensive historical ratings, provides a robust basis for testing these algorithms.

## Experiments

### Dataset

We used the Netflix dataset, containing explicit user ratings for movies, which is well-suited for evaluating recommendation algorithms.

### Model Configurations

1-SVD-based Model:

Learning Rate: 0.005

Regularization Term: 0.02

Number of Epochs: 5

2-Item-based Pearson Correlation:

Similarity Metric: Pearson correlation

User-based: False

**Evaluation Metrics**

RMSE (Root Mean Squared Error)

MAE (Mean Absolute Error)

**Results**

1-SVD-based Model:

Average RMSE: 0.890

Average MAE: 0.697

2-Item-based Pearson Correlation:

Average RMSE: 0.917

Average MAE: 0.726

## Discussion and Conclusion

Our experiments demonstrate that the SVD-based model outperforms the item-based Pearson correlation method in terms of predictive accuracy. This aligns with the hypothesis that SVD models, by capturing latent factors, provide better recommendations. The comprehensive evaluation, as guided by the FEVR framework, highlighted the complexity and multifaceted nature of recommender system evaluation. Future work could involve exploring hybrid models and incorporating additional user and item features to enhance performance further.

## References:

Zangerle, E., & Bauer, C. (2023). Evaluating Recommender Systems: Survey and Framework.

## Team Contribution:

Musaed owied  33% , Elyas Sindi 33%, Nawaf fayez alrehaili 33%