

FEVR

**Evaluating Recommender Systems:
Implementation and Analysis using
the Netflix Dataset**

Musaed owied – 2140988

Elyas Sindi – 2143512

Nawaf alrehaili -2142609

Table of CONTENTS

01

introduction

02

FEVR Framework

03

Netflix Dataset

04

Justification

05

Evaluation
Objectives

06

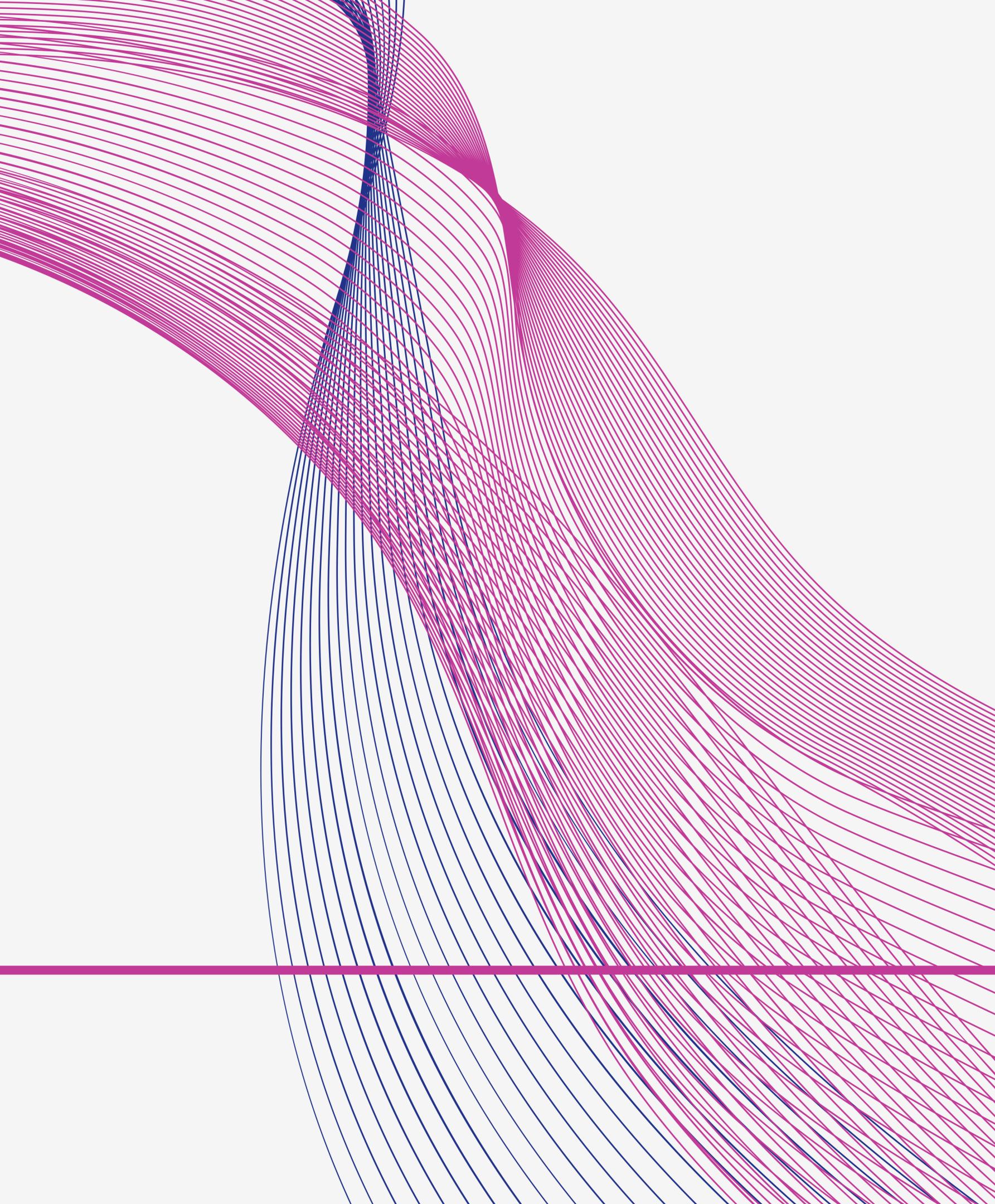
Evaluation
Principles

07

Evaluation Aspects

08

Result



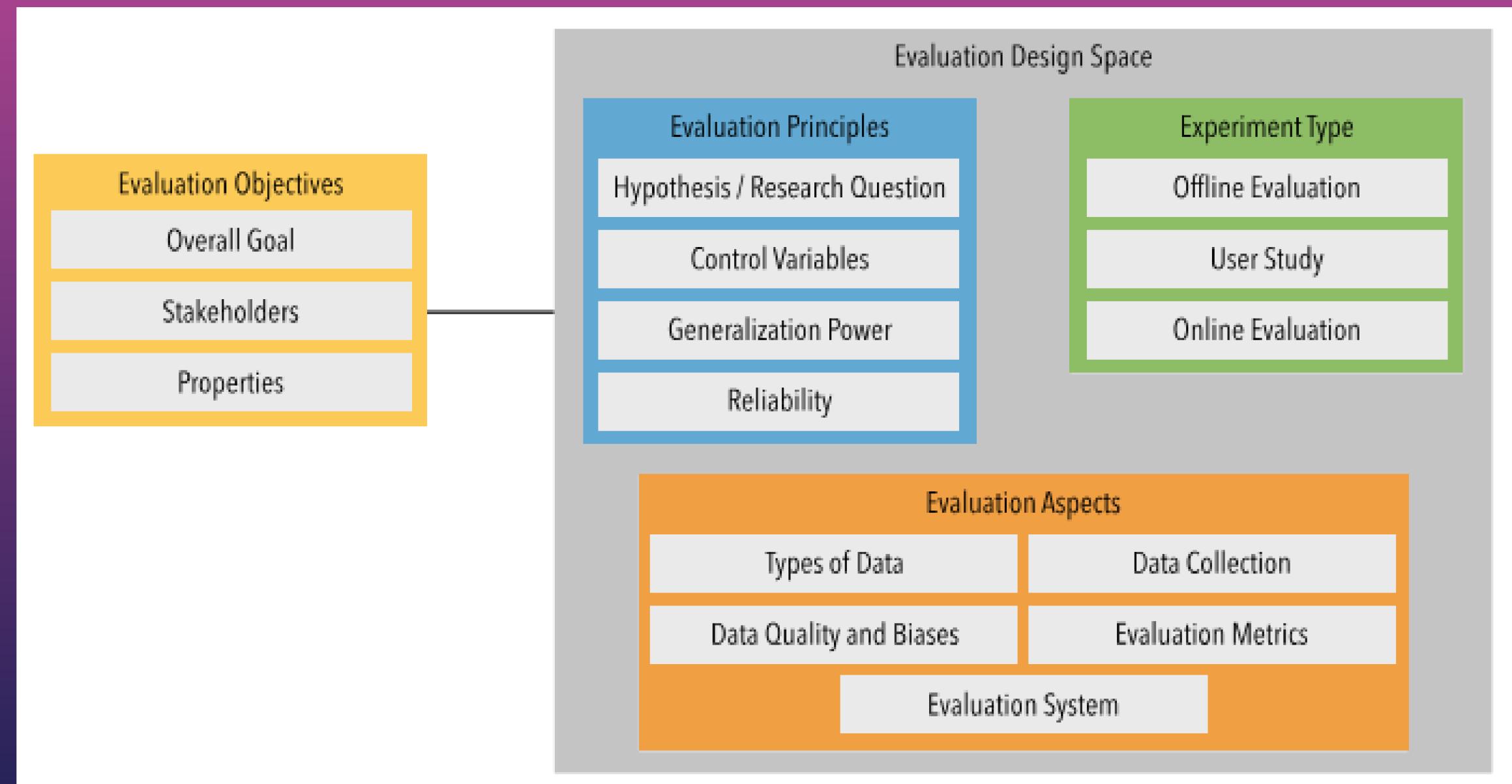
INTRODUCTION

- Implementation and evaluation of recommendation algorithms using the Netflix dataset.
 - Validate the FEVR framework's applicability for comprehensive RS evaluation.
 - Tested SVD-based models and item-based Pearson correlation methods.
-

FEVR Framework:

Framework for Evaluating Various Recommender systems

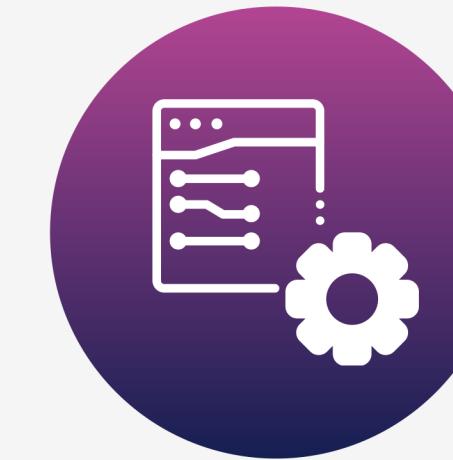
Summarizes and synthesizes existing literature to give readers a broad understanding of state-of-the-art, key developments, trends, and gaps in the research area.



NETFLIX DATASET



Contains explicit user ratings for a wide range of movies.



Widely used and well-documented, providing high-quality data for research

JUSTIFICATION



Alignment with FEVR Framework

Supports the comprehensive evaluation approach introduced in the research paper.



High-Quality Data

Well-documented and reliable data essential for accurate evaluation.



Extensive Historical Ratings

Facilitates robust offline evaluations of various recommendation algorithms.



Facilitates Hypothesis Testing

Enables testing of prediction accuracy of SVD-based models versus item-based Pearson correlation methods.

1

Evaluation Objectives

01

Prediction Accuracy

Objective:

Measure how accurately the recommender system predicts user ratings.

Metrics:

RMSE
MAE

02

User Satisfaction

Objective:

Assess user satisfaction with the recommendations.

Metrics:

User feedback: Gathered through surveys or user studies to determine user satisfaction levels. (Future Improvement)



2 Evaluation Principles

01

Hypotheses

- 1- SVD-based model provides higher prediction accuracy than Item-based Pearson Correlation.
- 2- Item-based Pearson Correlation method generates more diverse recommendations than SVD-based model.

03

Generalization Power

Evaluate models on different data subsets, use cross-validation, and test on various user groups or time periods.

02

Control Variables

Consistent setup, same dataset splits, preprocessing, metrics, and fixed random seed.

04

Reliability

Ensure consistent and reproducible evaluation, run multiple experiments and average results, document the process.

3

Experiment Types

01

Offline Evaluation

Use historical data, split into training and testing sets, train models, and measure performance on the test set.

03

Online Evaluation

Plan to deploy the model in a real-world setting and measure user interactions and satisfaction in a live environment. Plan to deploy the model in a real-world setting and measure user interactions and satisfaction in a live environment.
(Future Improvement)

02

User Study

Plan to gather qualitative feedback on user experience and satisfaction with the recommendations. (Future Improvement)



4 Evaluation Aspects

01

Types of Data

We should Use both explicit data (user ratings) and implicit data
In our case, we are using explicit ratings from the Netflix dataset.

02

Data Collection

We will use the existing Netflix dataset for offline evaluation.

03

Data Quality and Biases

Address potential biases in the data by using techniques like cross-validation and ensuring a representative sample.

04

Evaluation Metrics

RMSE
MAE

RESULT

MODELS	Average RMSE	MAE
SVD	0.890	0.697
Item-based Pearson Correlation	0.917	0.726

Conclusion

- **SVD Model Performance**
Outperforms Item-based Pearson Correlation in predictive accuracy.
- **Hypothesis Confirmation**
SVD models, by capturing latent factors, provide better recommendations.
- **Evaluation Insight**
FEVR framework highlights the complexity of recommender system evaluation.



THANKS