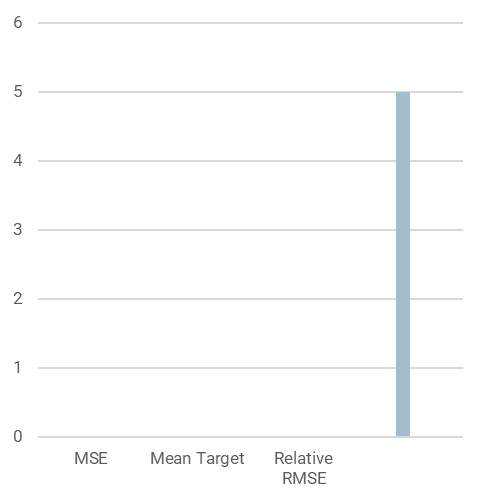
 ssssssssss



Methods

The NBA game history was turned into matrix representation. From that matrix were dropped random values to create restored matrix later. Were used as standard transformation to matrix representation and transformation of weighted dataset.

Through two databased data was united to include team stats from season 2018 to predict 2019 season results using CBF with RandomForestRegressor.

Later CBF and CF were combined into hybrid model using dynamic weights, Model Stacking, mean of two models and selectively use.

Results

There was no difference between CF on weighted dataset and pure dataset of games.

The results showed that the CF was greatly outperformed by CBF.

In case of hybrid recommendation systems, some of them were heavily affected by combining with CF, while others were somewhere in between CBF and CF performance (compared by relative RMSE). Model stacking was the only hybrid model, that has showed better results than CBF.

**Reference**

* *NBA stats (1947-present)*. (2024, December 1). Kaggle. <https://www.kaggle.com/datasets/sumitrodatta/nba-aba-baa-stats?resource=download&select=Advanced.csv>
* *NBA games data.* (2022, December 23). Kaggle. https://www.kaggle.com/datasets/nathanlauga/nba-games?select=players.csvss

Introduction

Predicting outcomes in complex sport domain such as NBA, requires accurate modeling of historical trends, team interaction and contextual features.

Traditional approaches like Collaborative Filtering (CF) excel in utilizing historical data, but often fail to adapt to new scenarios. On the other hand, Content-Based Filtering (CBF) leverage team attributes, but struggles with identifying patterns in historical dependencies.

Objective

* A collaborative filtering algorithm
* A content-based filtering algorithm
* A hybrid recommendation system combining the two algorithms by combining historical team matchup data (CF) and team-specific features (CBF)

NBA Team Score prediction using hybrid method of collaborative and content-based filtering

Team Performance Prediction using PLM

**Author**

Mikhail Karimov FIT CVUT

Analysis

CF showed strong performance, when sufficient historical interaction data between teams was available. However, relative RMSE for CF was higher in scenarios with limited interaction data or when CF reliance on historical pattern, which may not fully capture present-day context.

CBD demonstrated lower Relative RMSE in scenarios where contextual features played a significant role.

Weighted hybrid and selectively Use models were too much affected of not so accurate CF prediction. Mean prediction meantime, despite simplicity, outperformed all models, except CF and model stacking.

Conclusion

This study demonstrates the potential of the hybrid recommendation system in non-typical sport domain and comparation of CF, CBF and hybrid systems.