Final Project Part 4

The Random Forest model was selected for further hyperparameter tuning due to its strong feature interpretability and potential for improvement from its initial MSE of 1.96. I adjusted the genres by combining some of them to help our model overfit less. However, we saw that the model underperformed our expectations and achieved an MSE of 2.11. For the collaborative filtering approach, grid search identified KNNWithMeans with k=50 and Pearson similarity as the optimal configuration, achieving an RMSE of 1.43. Surprisingly, the simpler SVD matrix factorization outperformed this with a 1.39 RMSE, suggesting that this model may be better suited for this dataset. A hybrid approach combining the tuned Random Forest's content-based predictions (30% weight) with SVD's collaborative predictions (70% weight) achieved the best overall performance with a 1.32 RMSE, demonstrating the complementary strengths of both methods.

To further enhance the recommendation system, there are some ideas that may help. For example, developing a dynamic weighting mechanism for the hybrid approach could automatically adjust content vs. collaborative emphasis based on user history availability, and incorporating features like (rating timestamps, seasonal trends) could improve relevance. These enhancements would move the system from a pure prediction model to a more robust, production-ready recommendation engine.