## **EXECUTIVE REPORT**

## Executive Summary

FlickFlare, an innovative movie streaming platform, aims to enhance user satisfaction by introducing an advanced movie recommendation system. This system addresses user complaints about irrelevant movie suggestions by leveraging user ratings and genre preferences to recommend the top 5 movies tailored to individual tastes. Utilizing collaborative filtering methods and sophisticated algorithms such as k-Nearest Neighbours (KNN), Singular Value Decomposition (SVD), and Non-negative Matrix Factorization (NMF), we ensure highly accurate and relevant recommendations. Our analysis shows that NMF and SVD models significantly outperform the baseline, making them ideal for implementation. The primary goal is to enhance user interaction, improve recommendation accuracy, and ultimately increase user satisfaction and engagement.

## Introduction

FlickFlare offers a diverse and expansive collection of films, including classic cinema, indie gems, and the latest blockbusters. Despite the extensive selection, user feedback from the Google Play Store indicates dissatisfaction with the movie recommendations, which do not align with user interests. This has led to a decrease in user satisfaction and a 25% increase in churn rates over the past six months. To address this issue, FlickFlare is introducing a sophisticated movie recommender system that leverages user ratings and genre preferences to suggest the top 5 movies tailored to individual tastes.

## FlickFlare Movie Recommender System

## Overview

FlickFlare's recommendation system utilizes the MovieLens dataset to create a hybrid recommendation engine. This system provides personalized top 5 movie suggestions based on previous user ratings and preferences. By employing collaborative filtering techniques, FlickFlare ensures precise and relevant recommendations. The personalized approach aims to reduce content discovery friction, increase average watch time per user, and improve customer retention rates through a tailored viewing experience. This strategy helps FlickFlare differentiate itself from competitors by offering superior content curation and optimizing content acquisition and production strategies based on user preferences. This project aligns with FlickFlare's broader goal of leveraging data-driven insights to create a more compelling and personalized streaming service, driving business growth and solidifying its position in the competitive streaming industry.

## Business Understanding and Problem Statement

## FlickFlare faces a critical challenge:

- User feedback indicates dissatisfaction with movie recommendations.

- Current suggestions fail to align with user interests and preferences.

- This misalignment is impacting user engagement and retention, leading to a 25% increase in churn rates over the past six months.

To address this issue, FlickFlare has partnered with RODATA, a data analytics firm, to develop a sophisticated movie recommender system that provides personalized movie recommendations based on individual user ratings and preferred genres. This tailored solution aims to transform FlickFlare's recommendation engine, directly addressing user concerns and boosting overall platform performance.

## Objectives

## Main Objective

- To build a movie recommender system that suggests top movies to streaming users based on movie ratings.

## Specific Objectives

1. Implement and compare SVD, NMF, and KNN collaborative filtering algorithms for movie recommendation, evaluating their performance using RMSE on a benchmark dataset of user-movie ratings.

2. Use the model with the lowest RMSE to generate personalized top-N movie recommendations for streaming users.

3. Develop a hybrid recommender system that combines the strengths of SVD, NMF, and KNN models to potentially improve upon the performance of the best single model.

## Metric for Success

**Goal**: Achieve at least 80% accuracy in recommending the top five movies.

## Approach

**Model Implementation**: Utilize SVD, NMF, and KNN algorithms to predict user ratings and rank movies accordingly.

**Evaluation:** Measure the accuracy of predictions using RMSE. The goal is to achieve an RMSE that corresponds to at least 80% accuracy in the top five recommendations.

**Optimization:** Continuously tune and validate model parameters to improve accuracy. Techniques like cross-validation will be used to ensure the robustness of the models.

## Data Constraints

**Sparsity of the Data**: Many users rate only a few movies, resulting in a sparse ratings matrix. This makes it challenging to find commonalities between users or movies for collaborative filtering.

**Cold Start Problem**: Difficulty in recommending to new users or items with no rating history.

**Bias in Data**: Historical data may contain biases, such as popular movies receiving more ratings. These biases can skew recommendations towards already popular items, limiting diversity.

**Demographic Skew**: The dataset may not represent all user groups equally, leading to biased recommendations.

**Ensuring Balance:** It is crucial to balance accuracy and diversity in recommendations, offering precise suggestions while also providing a diverse range of options to avoid monotony and keep users engaged.

By addressing these challenges and objectives, FlickFlare aims to significantly improve its recommendation system, ensuring a more satisfying and engaging experience for its users.

## **Methodology**

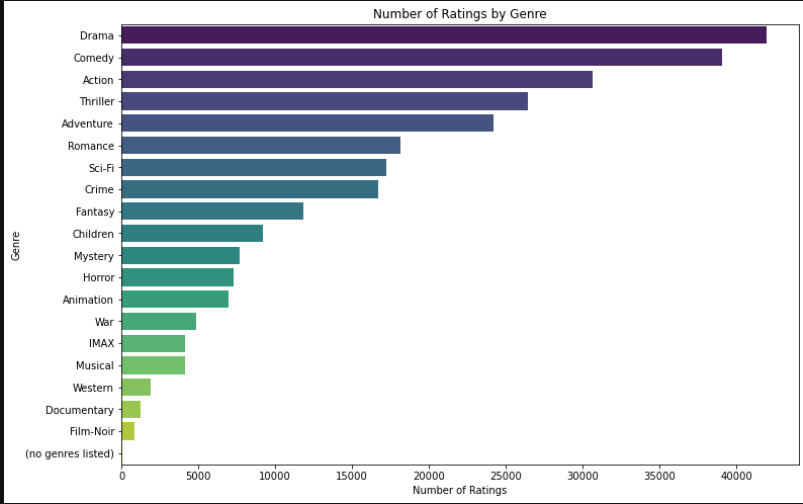
**In our analysis we used the CRISP-DM which stands for Cross-Industry Standard Process for Data Mining. It is a widely used methodology for data mining and analytics projects, providing a structured approach to planning and executing data mining tasks. It is composed of six major phases, each with specific tasks and deliverables: business understanding, data understanding, data Preparation, modelling and evaluation.**

**We used a technique called collaborative filtering to generate recommendations for users. This technique is based on the premise that similar people like similar things.**

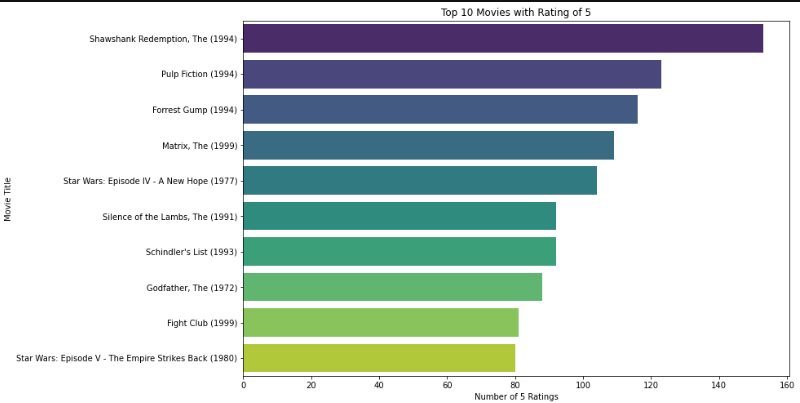
**When modelling our data to train it, we used k-Nearest Neighbours (KNN) to come up with a User-item Recommendations system. This will be our baseline model. A user-item recommendation system for a collaborative filtering movie recommender uses historical user interactions (like movie ratings) to suggest new movies. The core idea is that similar users tend to like similar items. KNN helps in identifying similar users by computing distances (similarities) in a high-dimensional space of user ratings. Once the nearest neighbours are found, the system predicts a user’s rating for a movie based on the ratings given by these neighbours. This method leverages the collective preferences of users, making it a powerful approach for personalized recommendations in a collaborative filtering system.**

**After establishing our baseline model and exploring the KNN algorithm, we decided to experiment with the Singular Value Decomposition (SVD) algorithm to enhance our model’s accuracy and reduce the RMSE. The surprise library implements a variant of SVD known as Funk's SVD, which relies on matrix factorization. We also use NMF (Non-negative Matrix Factorization) in collaborative filtering decomposes the user-item interaction matrix into two lower-dimensional matrices with non-negative values. This helps to identify latent features that represent user preferences and item characteristics, enabling the recommender system to predict missing ratings and make personalized recommendations based on the learned patterns.**

## **Key Findings**

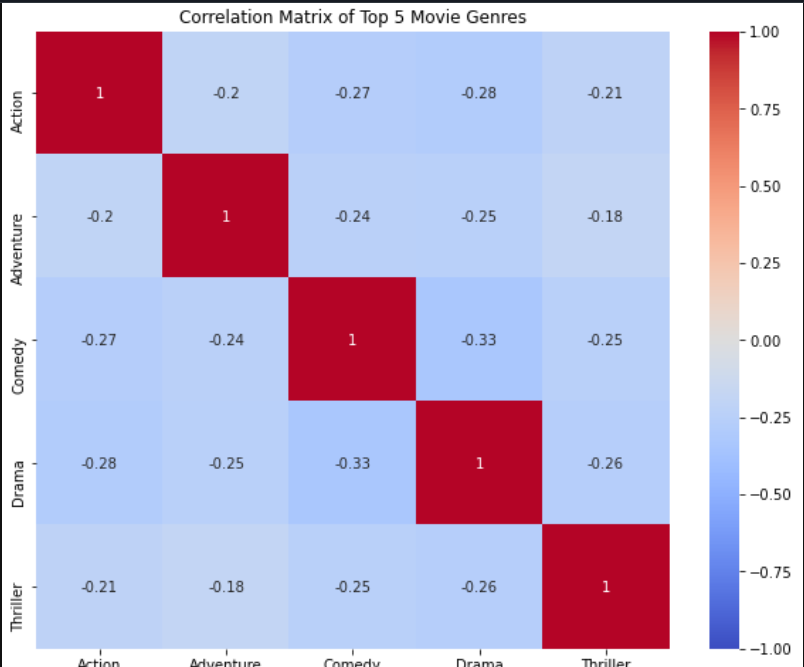
****

**We found that the movie categories that got the top 5 highest rating include: drama, comedy, action, thriller and adventure. This analysis will guide our recommendation as most people prefer these movie genres.**

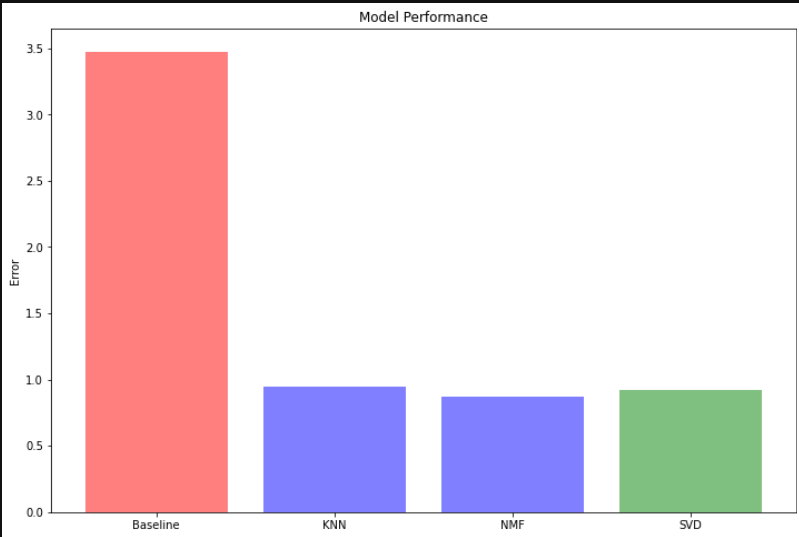
****

**This chart displays the top 10 movies that have received a perfect rating of 5, ranked by the number of such ratings they've received. "The Shawshank Redemption" (1994) has received the most 5-star ratings, with approximately 145 perfect scores. "Pulp Fiction" (1994) follows closely behind, with about 130 5-star ratings. The list includes a variety of genres, from drama (Shawshank Redemption, Schindler's List) to science fiction (The Matrix, Star Wars) to crime (Pulp Fiction, The Godfather).Two Star Wars movies appear on the list, showing the strength of this franchise among viewers who give perfect scores.**

**we use the correlation matrix to identify which genres are typically rated similarly by users. For example, if Action and Thriller have a high positive correlation, it might suggest that users who enjoy Action movies also enjoy Thrillers.**

****

**These results show very weak correlation between our movie genres which shows that our users have a variety of tastes. If a user one loves action movies it is not a guarantee they will automatically like comedy movies. This requires further analysis to develop a more personalized recommendation.**



**After training our data we observed that Model Performance: NMF and SVD models significantly reduced prediction errors by approximately 74% compared to the baseline.**

## **Recommendations**

* **We deploy the SVD model for the following reasons: Both NMF and SVD have the lowest RMSE, indicating they make the most accurate predictions among the models tested. The performance difference between NMF and SVD is minimal, so either could be a good choice. Both models show substantial improvement over the baseline, reducing error by about 74%. However, NMF appears to have a marginally lower RMSE than SVD, though the difference is small.**
* **For effectiveness of our movie recommendation system, we recommend implementing a feedback mechanism where users can rate the relevance of the recommendations they receive.**
* **Finally, it is crucial to ensure that the recommendation system is scalable to accommodate an increasing number of users and movie entries, potentially leveraging cloud-based solutions or distributed computing as necessary.**

## **Conclusion**

**The project has successfully addressed the core challenge of user dissatisfaction with movie recommendations, as previously noted in Google Play Store feedback. The system now successfully provides more personalized movie recommendations for users based on their past ratings.**

**By deploying the SVD model, FlickFlare can significantly improve the accuracy and relevance of movie recommendations, leading to enhanced user satisfaction and engagement**

## **Next Steps**

* **Combining Techniques: Integrate collaborative filtering with content-based filtering to leverage both user behaviour and item attributes. This can enhance the recommendation accuracy and address limitations of each individual approach.**
* **Cold Start Strategies: Develop methods to address the cold start issue for new users or items with insufficient data. Encourage new users to rate the movies they watch.**
* Incorporate user behaviour data such as watch history, search patterns, and time spent on different genres to further refine recommendations
* **By following these steps, FlickFlare can provide a more personalized and satisfying movie streaming experience, fostering greater user loyalty and engagement.**