A MOVIE RECOMMENDATION SYSTEM PROJECT

AGENDA

1. Introduction
2. Problem Statement
3. Objectives
4. Data Understanding & Preparation
5. Data Analysis
6. Modelling & Evaluation
7. Deployment
8. Conclusion,Recommendations & Next Steps

INTRODUCTION

Navigating the expansive film industry can be both exciting and overwhelming for consumers, given the diverse genres, directors, actors, and production styles. The surge in content on streaming platforms has made movie discovery challenging, prompting a need for personalized recommendations. To address this, a hybrid recommendation system that integrates collaborative filtering and content-based filtering is essential. This approach considers both user-item interactions and movie content features, ensuring a more tailored and efficient movie discovery experience for users.

PROBLEM STATEMENT

The modern film enthusiast faces a paradox of choice - a wealth of cinematic options, yet a struggle to find films that align with their preferences. The challenge lies not only in the initial selection but also in the subsequent quest for movies within the same niche or genre. Users often find themselves lost in the vast sea of content, seeking a solution that not only recommends the first movie but also facilitates a fluid journey through related titles.

OBJECTIVES

Enhance user satisfaction and engagement by delivering highly personalized and relevant movie recommendations

Improve customer retention by continuously tailoring suggestions based on changing user preferences

Increase active usage and interactions with the platform through accurate recommendations

DATA UNDERSTANDING AND PREPARATION

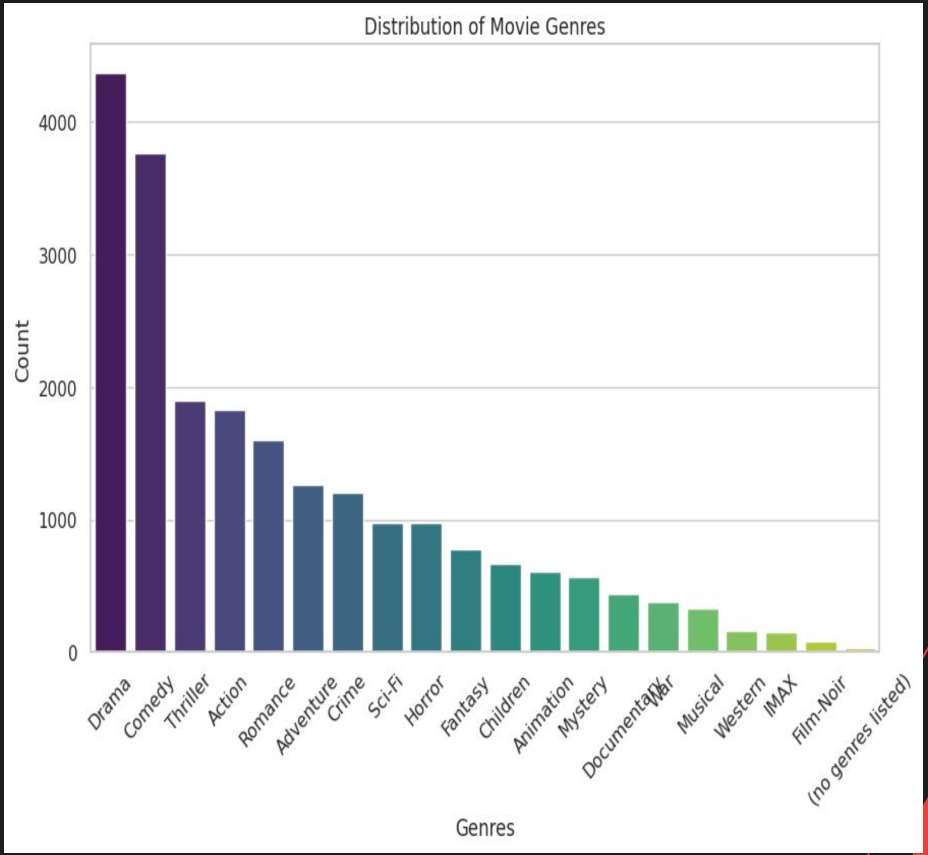
Our dataset, labeled *'ml-latest-small'*, is a comprehensive collection encompassing 100,836 ratings and 3,683 tag applications spread across 9,742 distinct movies. This rich dataset is the culmination of contributions from 610 individual users, spanning a period from March 29, 1996, to September 24, 2018.

A function is used open_csv(file_path), that reads a CSV file into a Pandas DataFrame. It then prints the first few rows, DataFrame information, handles null values, checks and drops duplicates, explores data distributions with histograms, and performs outlier detection using boxplots. The code also includes exception handling for potential errors during file reading.

DATA ANALYSIS

Bar plot showing distribution of movie genres.

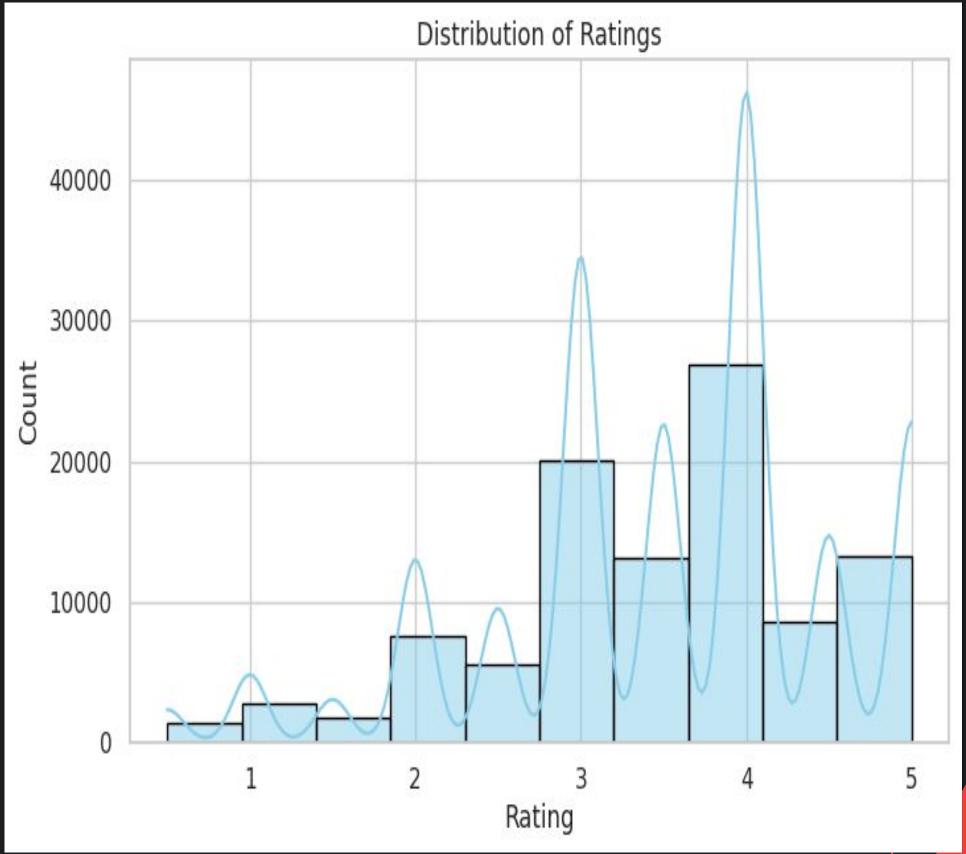
Most movies appear to fall under Drama and comedy.



DATA ANALYSIS

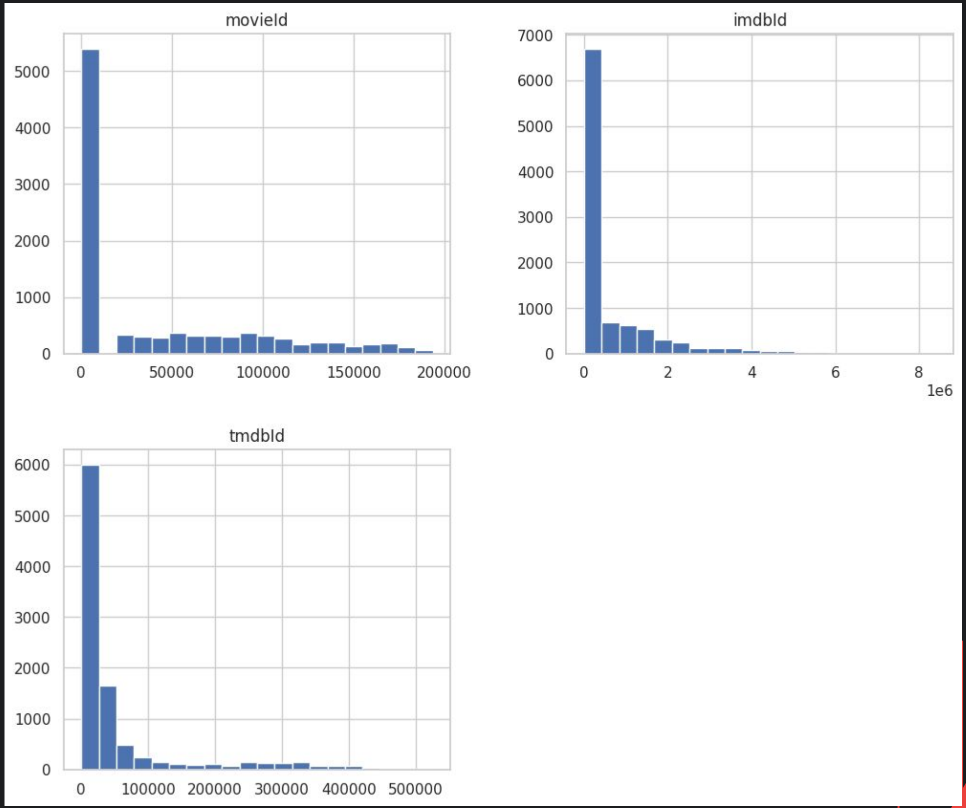
Histogram showing distribution of movie ratings.

Most users give a rating of between 3 and 4 indicating that most of them are satisfied with the movies.



DATA ANALYSIS

Histogram showing data distribution of various features



MODELLING AND EVALUATION

Popularity Based Recommender(Baseline model)

This is the easiest type of recommendation system based on item popularity.

It is also user cold start resistant

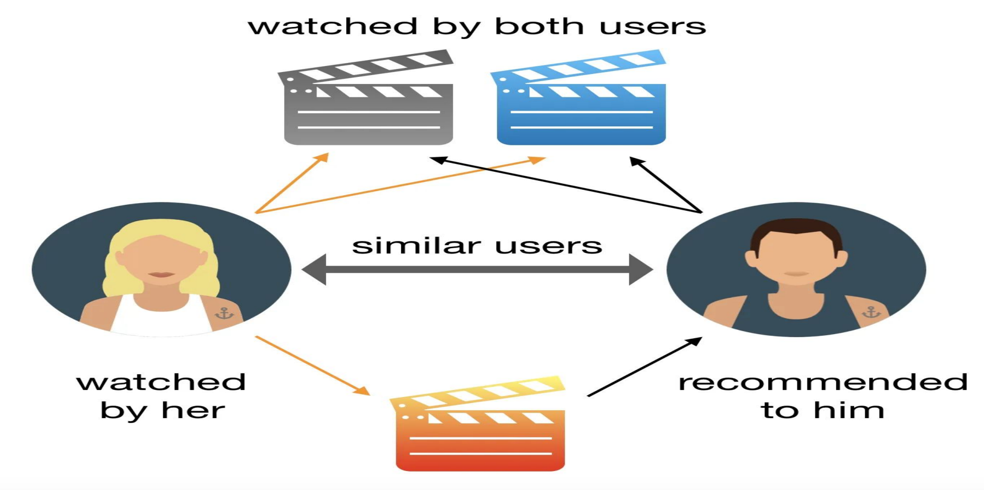
The Popularity-Based Recommender's RMSE of approximately 1.0425 indicates a substantial deviation, considering the rating scale of 0.5 to 5

This high RMSE highlights the need for a more effective model with improved accuracy.

MODELLING AND EVALUATION

Advanced Modelling

This segment introduces User-Based Collaborative Filtering, Singular Value Decomposition (SVD), K-Nearest Neighbors (KNN) with means, and K-Nearest Neighbors (KNN) basic (KNNBasic) models, along with their respective RMSE values.



Results

User-Based Collaborative Filtering RMSE: 0.9508204907011504

K-Nearest Neighbors with Means RMSE: 0.8999256300443209

SVD RMSE: 0.8808760139456194

K-Nearest Neighbors Basic RMSE: 0.9133878300866659

The SVD model has the lowest RMSE, meaning it is the most accurate in predicting movie ratings on unseen data

DEPLOYMENT

To bring this system to life, we chose Streamlit for its deployment.

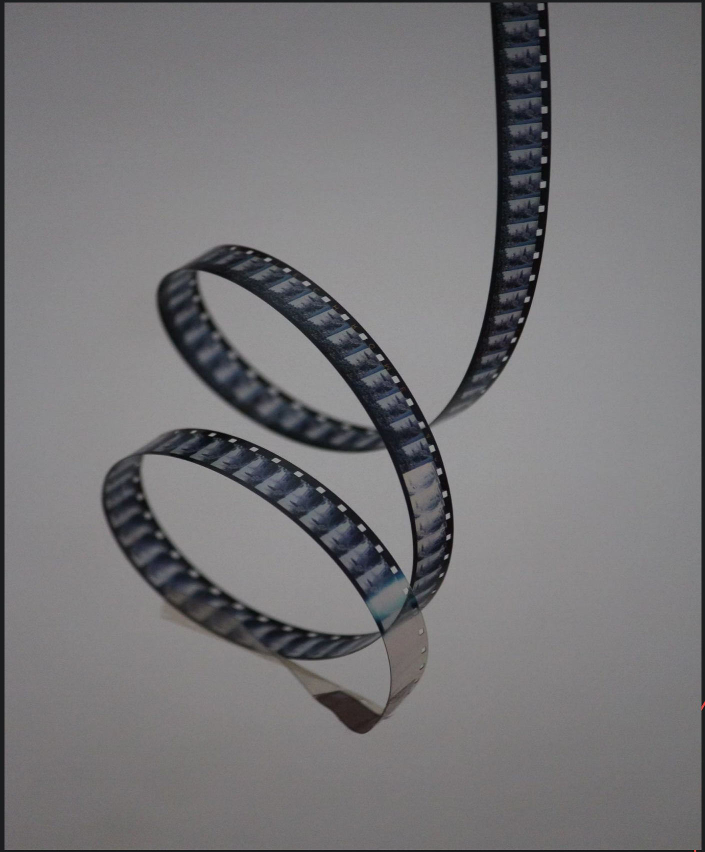
The platform allows us to demonstrate the system's capabilities in an interactive and user-friendly interface.

CONCLUSIONS

The recommendation system boasts an impressive 86% accuracy after tuning, easing content navigation and enhancing user satisfaction.

By streamlining movie searches and delivering personalized recommendations, it maximizes content enjoyment, fostering efficiency, and cultivating user loyalty.

This translates to prolonged user engagement, positively impacting long-term retention and platform prosperity.



RECOMMENDATIONS

1. Explore the development of a hybrid recommender system, combining the strengths of the SVD model and a content-based approach. This integration aims to maximize the benefits of both methods for an enhanced user recommendation system.

2. Introduce content-based recommendation features, leveraging the analysis of movie attributes like genre, actors, directors, and individual user preferences for a more varied and personalized recommendation experience.

3. Opt for showcasing films with a minimum rating of 3.5 and above, as these tend to appeal to a broad user base.

NEXT STEPS

1.Implement content-based recommendations analyzing attributes like genre, actors, and directors.

2. Develop a hybrid recommender system, combining SVD model and content-based approaches.

3. Establish a feedback loop for continuous algorithm refinement based on user preferences and engagement metrics.

THE TEAM

KIPKOSGEI
KIPTUI

LEWIS
KAMINDU

GLORIA
NYONGESA

CYNTHIA
KAREKO

PAUL
KAMAU

SHUAIB
MAHAMUD

THANK YOU!