# Movie Recommendation System: A Personalized Approach Using Weighted Ratings and Linear Regression

Meghana D. R. [1], Umm E Asma [2], Asma Noorain [3], Sindu N [4]

*Assistant Professor, PES Institute of Advanced Management Studies, Shivamogga*

*Abstract*—**With the exponential growth of online streaming platforms, choosing a movie that aligns with user preferences has become a challenging task due to the vast amount of content available. This paper presents a Movie Recommendation System that enhances the recommendation process by considering not only IMDb ratings but also the number of user votes. We introduce a weighted rating model, which improves the accuracy of the recommendations by factoring in both the quality and popularity of the movies. Additionally, a Linear Regression model is employed to predict the revenue of a movie based on its rating. The system utilizes Python libraries such as Pandas, Matplotlib, and Scikit-Learn for data processing, visualization, and predictive modelling. The experimental results demonstrate the effectiveness of the system in providing personalized movie recommendations and revenue predictions.**

## 1. INTRODUCTION

Movies have become an essential part of our daily entertainment, providing relaxation and recreation to individuals worldwide. However, the sheer volume of available movies makes it difficult for users to choose a movie based on their preferences. To address this issue, movie recommendation systems have become an indispensable tool for streaming services. These systems help users navigate through a wide range of movies by offering personalized suggestions.

Traditional recommendation systems typically rely on a single rating or popularity-based filtering. However, this approach may not accurately reflect the movie's true value, especially when movies have a large number of ratings but relatively few user votes. In this research, we propose a **Movie Recommendation System** that uses **weighted ratings**, incorporating both the IMDb rating and the number of user votes, to offer more accurate recommendations. Additionally, we utilize a **Linear Regression model** to predict the revenue of a movie based on its ratings, offering a deeper understanding of the factors influencing movie success.

## 2. MOTIVATION AND OBJECTIVES

### 2.1 Motivation

The growing availability of movie-related data on the web poses an information overload challenge for users. The ability to personalize the movie selection process based on individual preferences is critical to improving user experience. The motivation behind this research is to enhance traditional recommendation systems by incorporating a weighted rating system and predicting movie revenues using Linear Regression. This approach aims to provide users with more relevant and insightful recommendations.

### 2.2 Objectives

The objectives of this research are as follows:

1. To develop a weighted rating system based on IMDb ratings and the number of user votes to improve the accuracy of movie recommendations.
2. To provide personalized movie recommendations based on user preferences and weighted ratings.
3. To analyse genre distribution and movie revenue predictions, offering users insights into movie trends.
4. To implement a Linear Regression model to predict movie revenue based on its rating, contributing to a more informed recommendation process.

## 3. RELATED WORK

Recommendation systems have been widely researched, particularly in the context of e-commerce, entertainment, and content-based filtering.

- Konstan and Riedl (2012) reviewed the evolution of recommendation algorithms and the

importance of personalized user experiences in recommender systems.

- Rashid et al. (2002) proposed adaptive systems that learn user preferences over time, improving the quality of recommendations in dynamic environments.
- Schafer et al. (1999) introduced collaborative filtering techniques, which laid the foundation for modern recommender systems.

The common theme across these studies is the importance of personalization and adaptive learning, which are central to the Movie Recommendation System proposed in this paper.

## 4. METHODOLOGY

The proposed Movie Recommendation System follows a multi-phase approach, incorporating data collection, data cleaning, manipulation, and predictive modelling. The workflow is illustrated below:

4.1 Data Collection

Two primary datasets, Movies.csv and Ratings.csv, are utilized in the project. These datasets contain movie titles, genres, revenues, ratings, and the number of user votes.

4.2 Data Preprocessing and Cleaning

To ensure the quality of the data, any missing or null values in the datasets are handled by filling in appropriate default values. The merge () function from Pandas is used to combine the two datasets based on a common column (Movie ID).

4.3 Weighted Rating Calculation

To compute the weighted rating for each movie, we use the formula:

Weighted Rating=(Number of VotesNumber of Votes +Minimum Votes×Average Rating)+(Minimum Votes Number of Votes+Minimum Votes×Mean Rating)\text{Weighted Rating} = \left( \frac{\text{Number of Votes}}{\text{Number of Votes} + \text{Minimum Votes}} \times \text{Average Rating} \right) + \left( \frac{\text{Minimum Votes}}{\text{Number of Votes} + \text{Minimum Votes}} \times \text{Mean Rating}

\right)Weighted Rating=(Number of Votes+Minimum VotesNumber of Votes ×Average Rating)+(Number of Votes+Minimum Votes Minimum Votes×Mean Rating)

This formula combines both the average rating and the popularity (vote count) of each movie to generate a more accurate weighted score.

4.4 Linear Regression Model for Revenue Prediction

A Linear Regression model is employed to predict the revenue of a movie based on its IMDb rating. The model is trained using the formula:

Revenue=β0+β1×Rating\text{Revenue} = \beta_0 + \beta_1 \times \text{Rating}Revenue=β0+β1×Rating

This regression analysis allows for better understanding of the relationship between ratings and revenue, and provides valuable insights for users when making decisions.
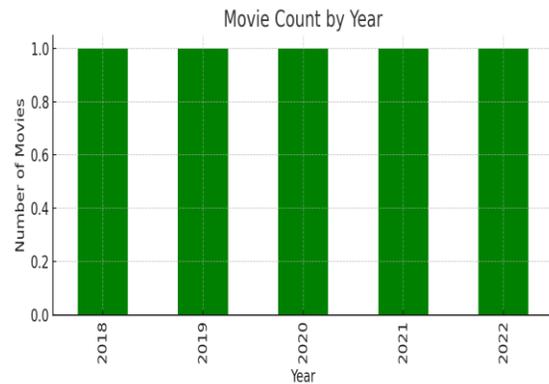
4.5 Data Visualization

Various data visualization techniques, including line graphs, bar charts, scatter plots, and pie charts, are employed to present the findings in an insightful manner. For example:
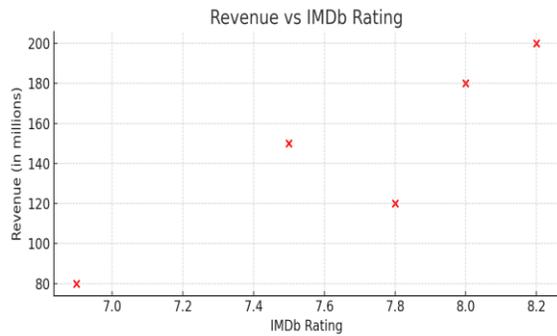
- A line graph shows the relationship between movie titles and ratings.
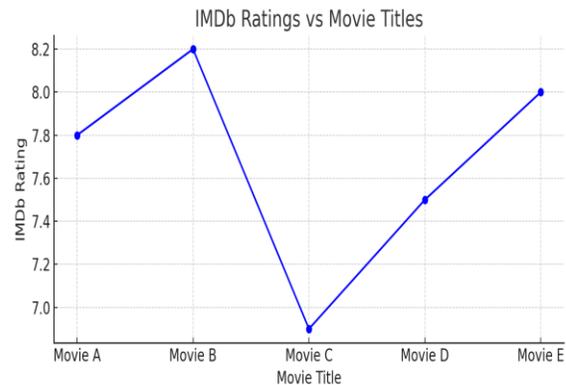


- A bar chart displays the number of movies released in each year.



- A scatter plot visualizes the correlation between movie ratings and revenue.

Revenue vs IMDb Rating

## 5. ADVANCED FEATURES

To enhance the system's utility, the following advanced features have been incorporated:

5.1 Genre Prediction
The system allows users to filter movie recommendations by genre. The most popular genres in the top movie list are identified and presented in a pie chart, helping users choose movies based on their preferred genres.

5.2 Box Office Forecasting
The system predicts a movie's box office revenue using the Linear Regression model. This feature can guide users in selecting movies with high commercial potential.

5.3 Sentiment Analysis for Reviews
To further improve recommendations, we propose incorporating sentiment analysis of user reviews. By analysing the sentiment of reviews, the system can provide a more nuanced recommendation by considering both the movie's rating and user sentiment.

## 6. RESULTS AND DISCUSSION

6.1 Movie Recommendations
The system generates a list of **top movies** based on the weighted ratings, allowing users to select the number of top movies they wish to view. The results are displayed with movie titles, ratings, and weighted ratings.

6.2 Genre and Year Distribution
The bar chart shows the distribution of movies in different genres and the number of movies released in each year, which provides insight into movie trends.

6.3 Revenue Prediction
By applying the Linear Regression model, the system accurately predicts the revenue of a movie based on its

rating. This prediction is useful for users seeking high-grossing films.

## 7. CONCLUSION

The proposed Movie Recommendation System enhances traditional recommendation models by incorporating weighted ratings and predicting movie revenue. This system provides users with more personalized recommendations by considering both the quantity and quality of ratings. The integration of advanced features such as genre prediction and revenue forecasting further improves the system's utility. Future work may involve incorporating sentiment analysis and deep learning models to refine predictions and offer even more accurate movie recommendations.

## REFERENCES

[1] Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. User Modeling and User-Adapted Interaction, 22(2), 101-124.

[2] Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., & Riedl, J. (2002). Getting to Know You: Learning New User Preferences in Recommender Systems. In: Proceedings of the International Conference on Intelligent User Interfaces, 127-134.

[3] Schafer, J. B., Konstan, J. A., & Riedl, J. (1999). Recommender Systems in E-Commerce. In: Proceedings of the 1st ACM Conference on Electronic Commerce, 158-167.

[4] Pan, C., & Li, W. (2010). Research Paper Recommendation with Topic Analysis. IEEE Computer Design and Applications, 8(3), 332-337.

[5] Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2014). Empirical Analysis of the Impact of Recommender Systems on Sales. Journal of Management Information Systems, 30(1), 179-204.