Brandon Mather

Personalized Movie Recommendation Model

Summary:

In the evolving landscape of entertainment, the ability to provide personalized movie recommendations can become a key differentiator for streaming platforms seeking to enhance user satisfaction and engagement. This white paper will show the developments of a movie recommendation system using predictive modeling. By embracing these techniques streaming platforms can create new opportunities for better movie recommendations that resonate with user preferences.

Business Problem:

Streaming platforms are always striving to enhance user experience by providing personalized recommendations that are based on individual preferences. The ability to accurately predict user preferences and suggest relevant content not only improves user satisfaction but also increases engagements and retention. Developing an effective movie recommendation system is an important need for streaming platforms seeking to gain a competitive advantage in the market.

History:

Movie recommendation systems have evolved over the years, driven by advancements in machine learning and data analytics. Traditional approaches relied on simple algorithms such as popularity-based or genre-based recommendations. However, with the rise of predictive modeling, recommendations systems have become more sophisticated and capable of providing even better suggestions based on user behavior and preferences.

Data Explanation:

The dataset that is being used is from TMDb (The Movie Database) and provides information about over 1,000,000 movies. The information includes titles, ratings, release dates, revenue, genres, and more.

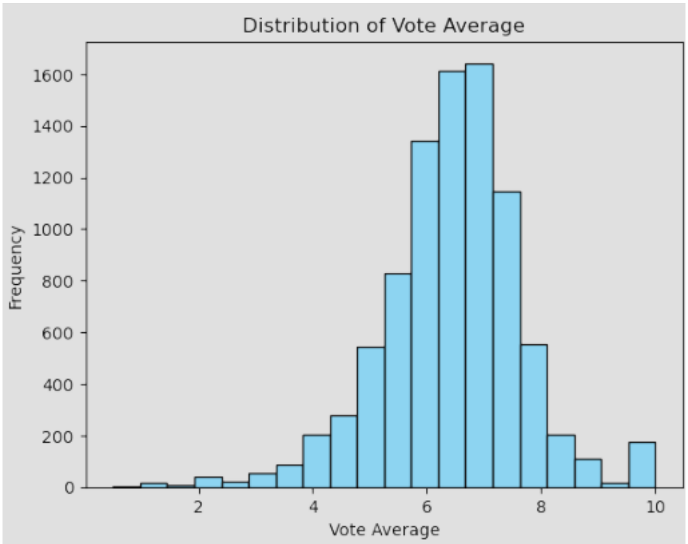
Methods:

The methods used were a combination of data preprocessing, exploratory data analysis (EDA), and predictive modeling. Preprocessing was used to make sure the data was usable and clean, the EDA revealed insights into the data, and predictive modeling was used to forecast movie recommendations.

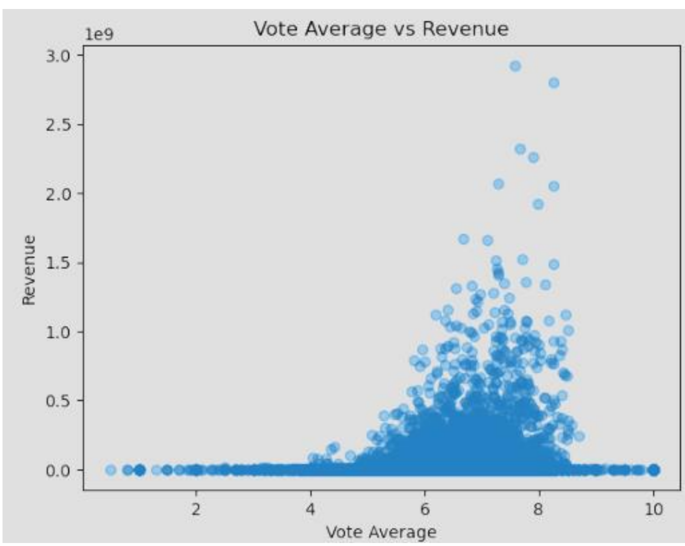
Analysis:

The first step in my analysis was to preprocess my data, and make sure it was usable. The first thing I did was get rid of any movie that had no score. These movies as of right now wouldn't be any help to the recommendation model because there is no history of it being watched. The second cleaning step was to convert the release date column to datetime so the model could read that column. The last step for the preprocessing was getting rid of any blank values.

Once the data was cleaned, I wanted to do a EDA (Exploratory Data Analysis) to visualize my data. Here we can see where the Vote Average distribution of every movie is, which tells us that most movies fall within the 5-8 range in likability:

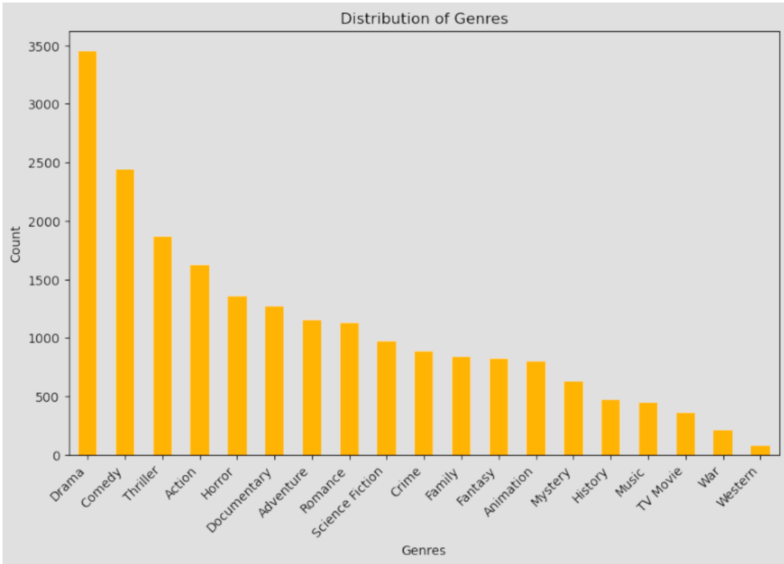


Then I wanted to see if revenue had any impact on the vote average:



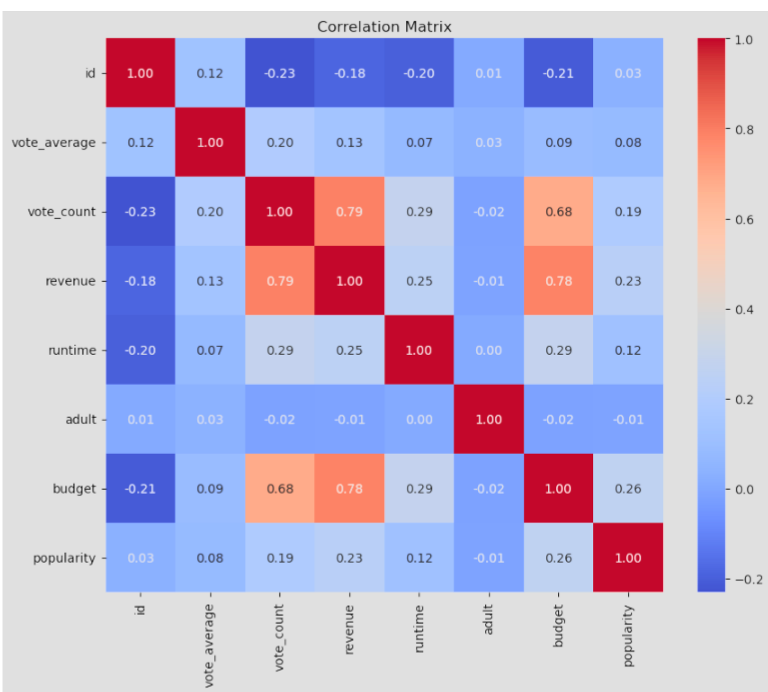
Here we can see that the amount of revenue spent on a movie can have a positive impact on the ratings, but not a guarantee.

I also want to make sure my movie dataset has a somewhat equal distribution of movie genres:



In this bar graph, it does show that drama has the most movies, by a good margin. Most of the other genres are closer, so this could be showing that dramas are much more popular than other movie genres, but we want to make sure the model is biased towards drama movies.

The last thing I wanted to see was the correlation between all the columns, which could give good insight:



After visualizing the data, I created the prediction model using scikit's Surprise, which is specifically used for building and analyzing recommender systems with rating data. Using movies watched by the user the model will try to predict the ratings for movies not watched to help give the highest predicted ratings as the recommendations. The model was evaluated using Root Mean Squared Error (RMSE) and was tested to create a recommendation.

Conclusion:

The analysis shows how predictive modeling can help with movie recommendations. By leveraging these techniques, streaming platforms can enhance user experience, increase user engagement, and drive business growth.

Assumptions:

The biggest assumptions that are made are that the ratings provided by users in the dataset accurately reflect their real preferences and users' preferences are consistent over time. Another assumption is that the dataset contains sufficient information to train a robust recommendation model.

Limitations:

Limitations include that the model may struggle at the start for a new user, since there will be limited information about their preferences. Movies with limited ratings also may struggle at the start. The model relies on the availability of user interaction data, which may be sparse in some cases. The performance heavily depends on the quality and quantity of the training data as well.

Challenges:

Challenges that may present themselves are scalability issues when dealing with large datasets and high user traffic, mitigating biases in the recommendation system, and addressing privacy concerns as it relates to the collection and use of user data for the recommendations.

Future Uses:

Future uses could explore a hybrid recommendation system using both predictive modeling and collaborative filtering techniques. Also, we could integrate real-time user feedback to adapt recommendations dynamically based on user interactions.

Recommendations:

I would recommend continuously refining and optimizing the recommendation model based on user feedback, implementing A/B testing to evaluate the impact of recommendation models on user engagement and retention, and create transparency with the user and gain their trust by providing clear explanations of how recommendations are generated and ensuring data privacy and security.

Implementation Plan:

The implementation plan would be first to gather user ratings data and preprocess it for modeling, then train the prediction model, assess the performance of the recommendation model using appropriate metrics. Then integrate the recommendation model into the streaming platform's user interface and backend infrastructure, and continuously monitor the model's performance, gather user feedback, and make necessary adjustments to improve accuracy.

Ethical Assessment:

We would make sure to ensure compliance with data privacy regulations and obtain user consent for collecting and using their data for recommendations, mitigate biases in the model to ensure fair treatment to all users, provide clear explanations on how recommendations are generated and allow user to control their preferences and privacy settings, and establish mechanisms for accountability and oversight to address and ethical concerns of the model.

References:

"Full TMDB Movies Dataset 2023 (930K Movies)." *www.kaggle.com,* www.kaggle.com/datasets/asaniczka/tmdb-movies-dataset-2023-930k-movies.

10 Questions:

1. How does the prediction model for recommendations compare to other recommendation approaches?

   ANSWER:

2. What measures were taken to ensure the accuracy and reliability of the recommendation model?

   ANSWER:

3. How do you handle new users or movies with limited ratings in the recommendation system?

   ANSWER:

4. Could you elaborate on the scalability challenges faced when implementing the recommendation model?

   ANSWER:

5. What strategies are in place to address biases?

   ANSWER:

6. How does the model handle user privacy and data security, and what measures are taken to ensure compliance?

   ANSWER:

7. Can you provide examples of how user feedback and interactions are incorporated into the model?

<u>ANSWER</u>:

8. What are some unintended consequences associated with deploying a recommendation model?

<u>ANSWER</u>:

9. What are best practices for implementing and maintaining the recommendation model in a real-world setting?

<u>ANSWER</u>:

10. Do you have any unanswered questions that warrant further investigation?

<u>ANSWER</u>: