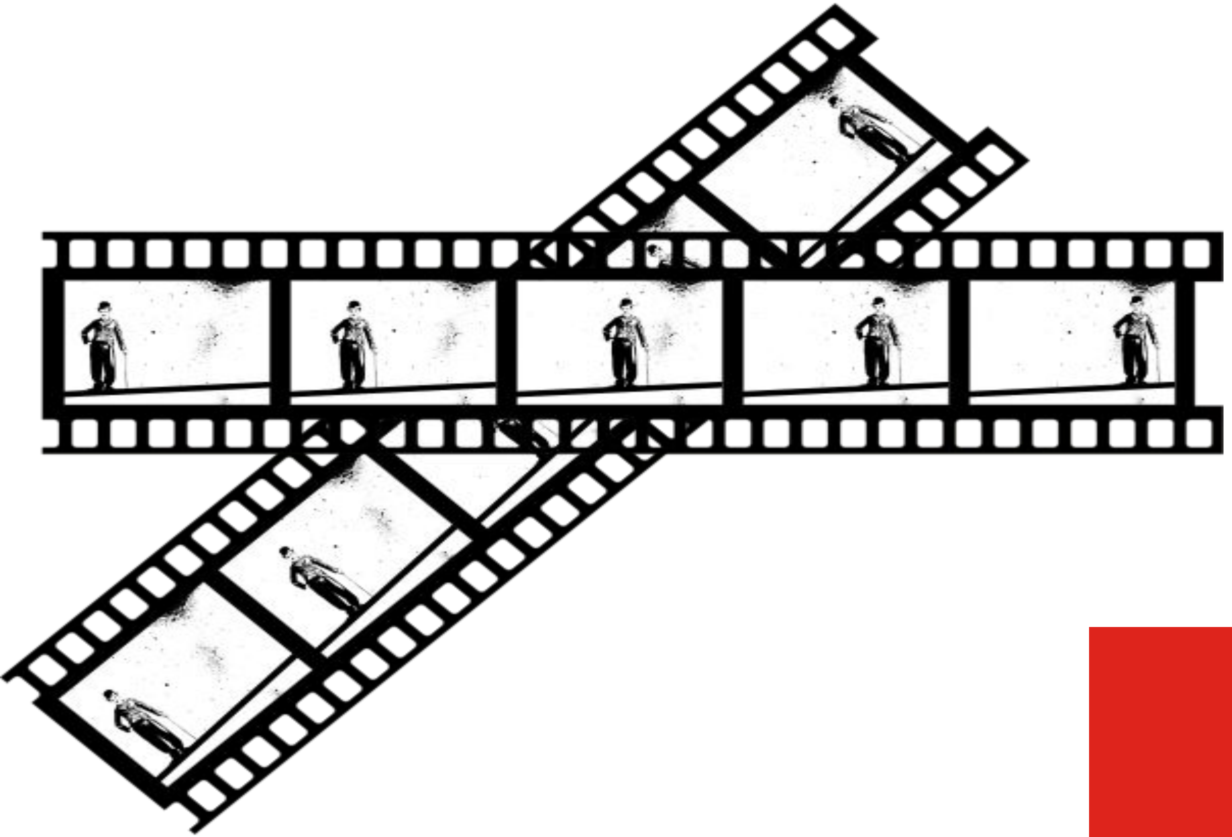
Charlotte's Movie Recommendation System

BY Charlotte Sun

2024.10.1

Overview



01 | Exploratory Data Analysis

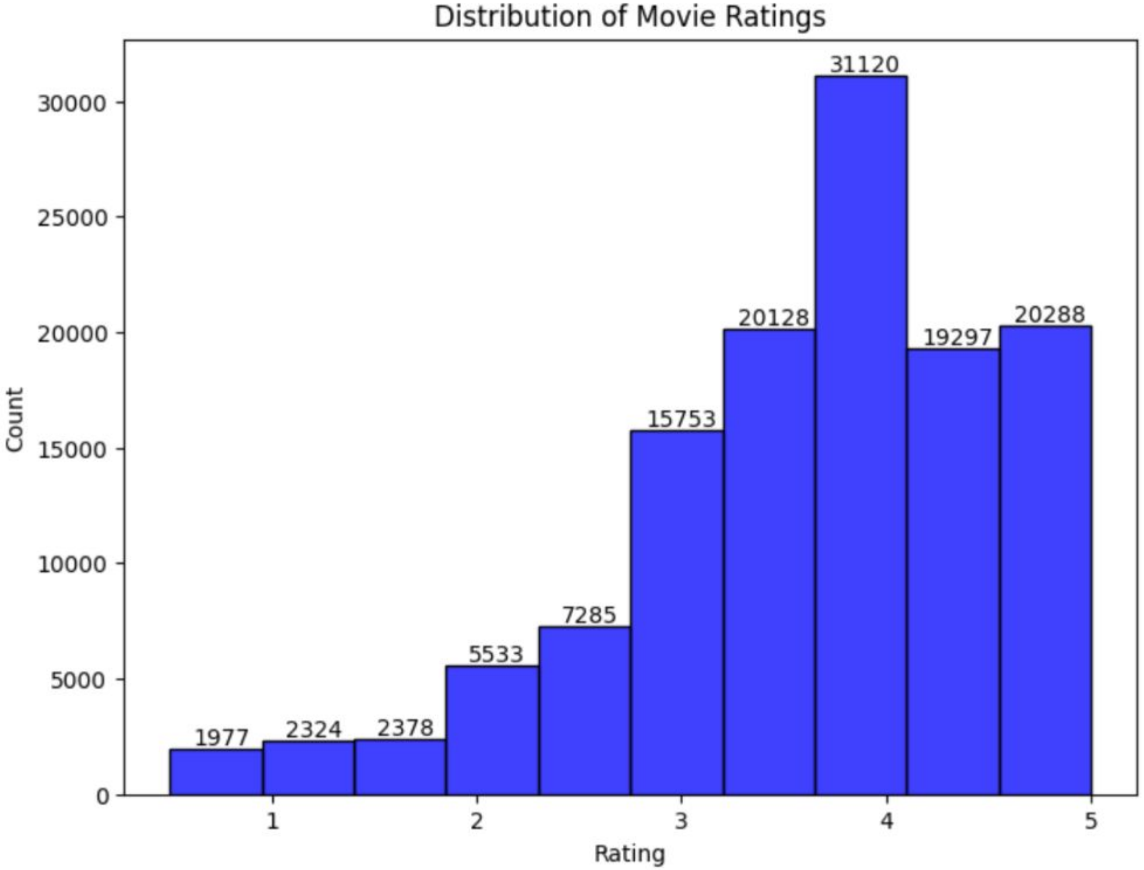
02 | Collaborative Filtering

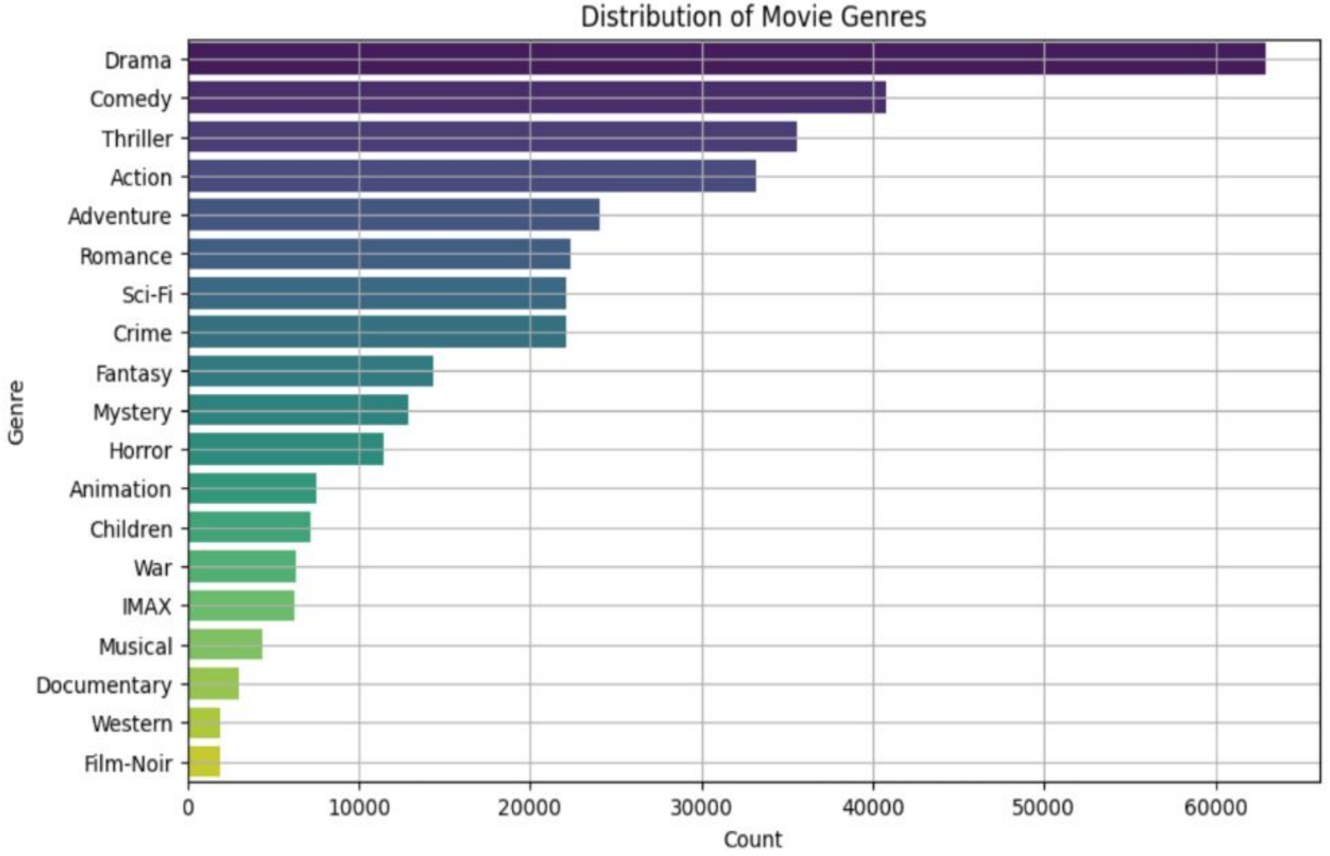
03 | Neural Network Structure

04 | Model Performance & Result

05 | Transformer-Based Recommendation System

Exploratory Data Analysis





- The rating distribution is positively skewed.
- This trend suggests that users generally rate movies they enjoy rather than those they dislike.

- Drama is the most common genre, followed by Comedy, Thriller, and Action.
- Less popular genres such as Film-Noir, Western, and Musical have significantly fewer movies.

Collaborative Filtering

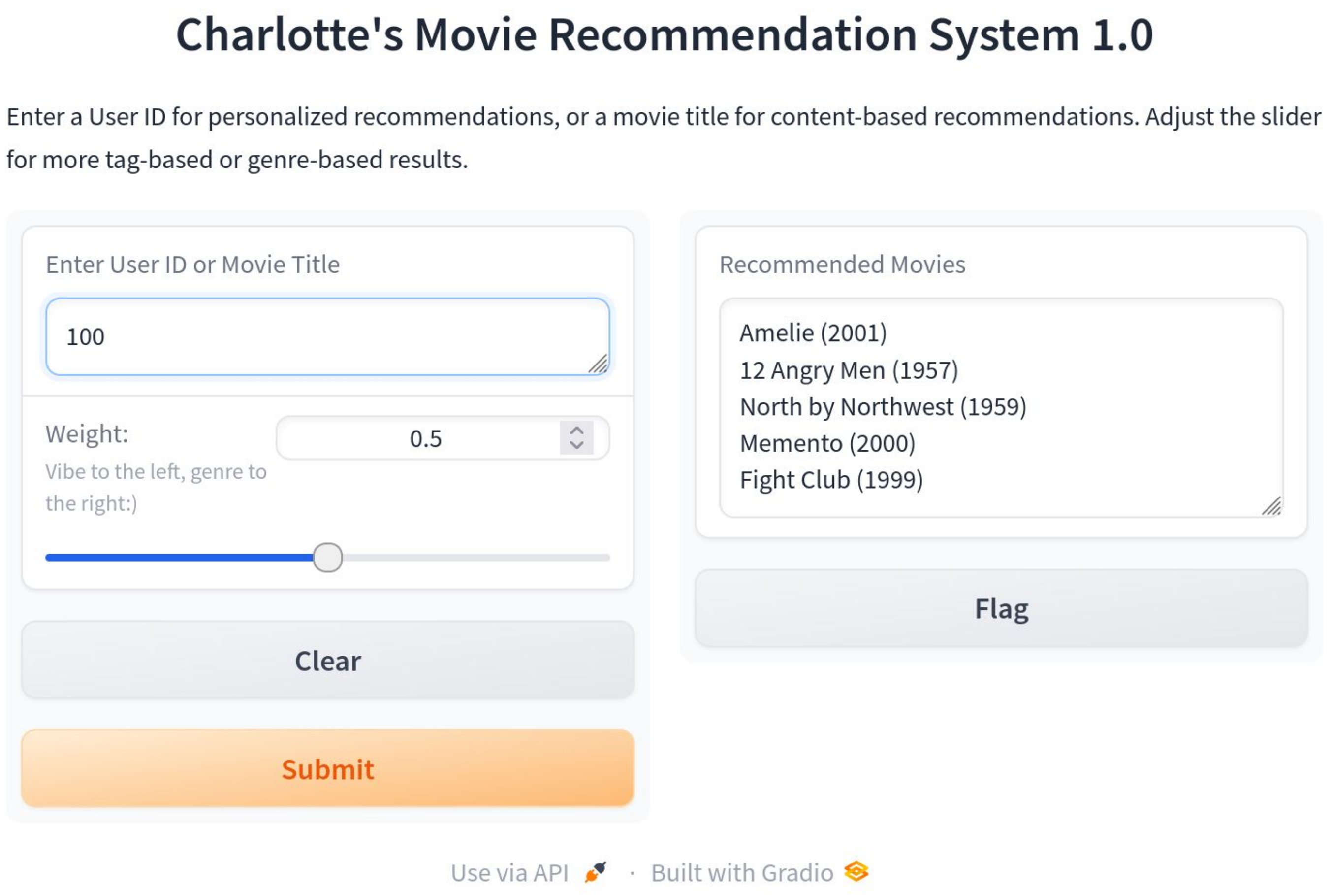
Data Preparation:
Convert user ratings into a sparse matrix for user-movie interactions.

Filtering the Data:
Focus on active users and popular movies by filtering out low-frequency interactions.

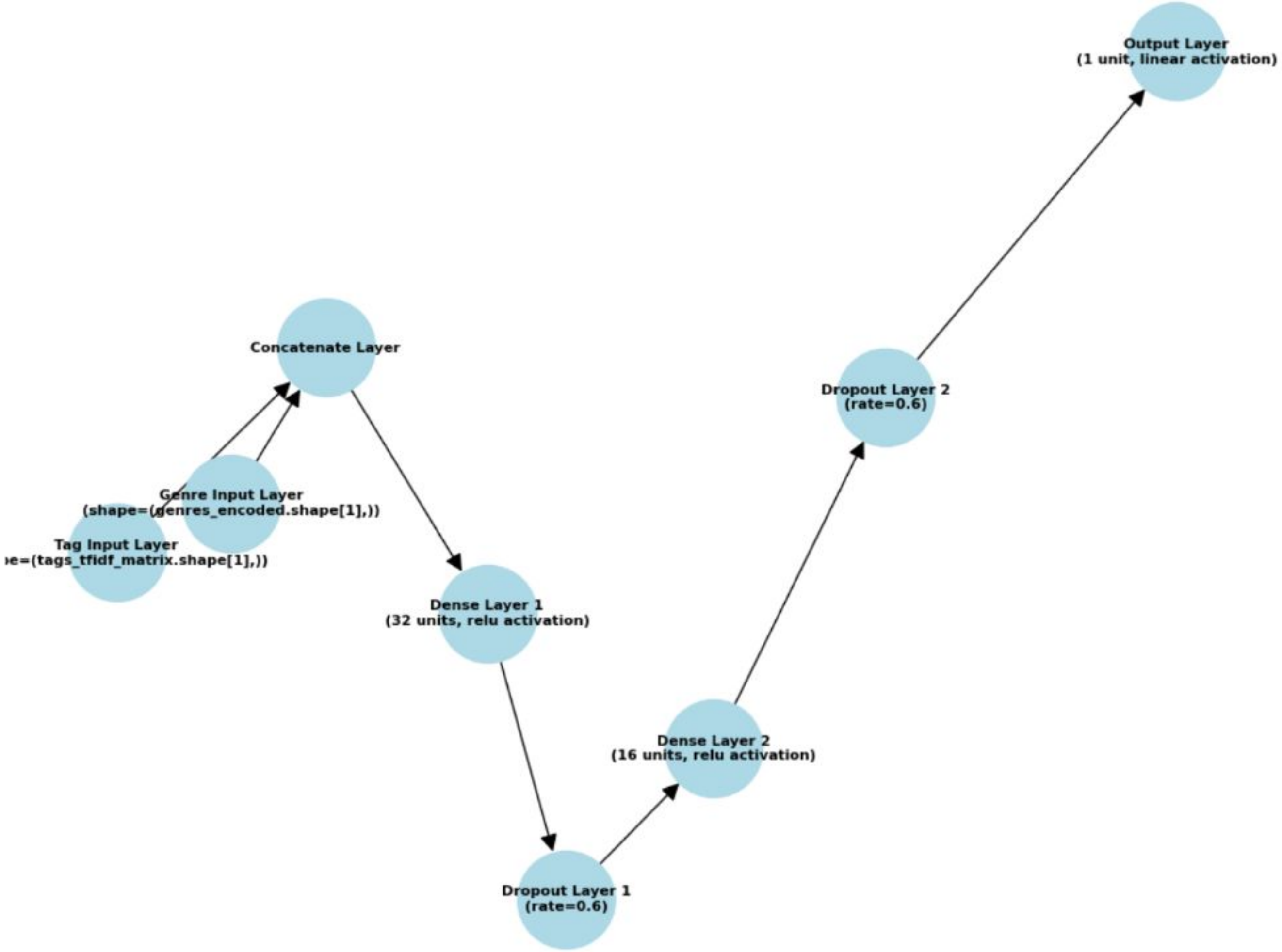
Label Encoding:
Encode user and movie IDs into numerical labels for efficient computation.

SVD Model Training:
Decompose the rating matrix using SVD and fine-tune parameters with GridSearchCV.

Recommendation Generation:
Predict ratings for unrated movies and recommend top movies to the user.



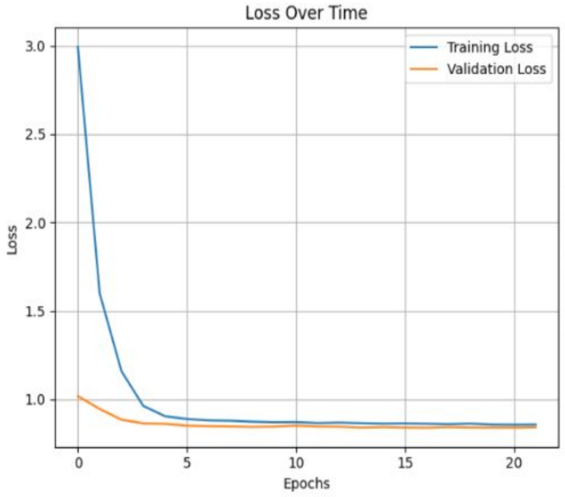
Neural Network Structure

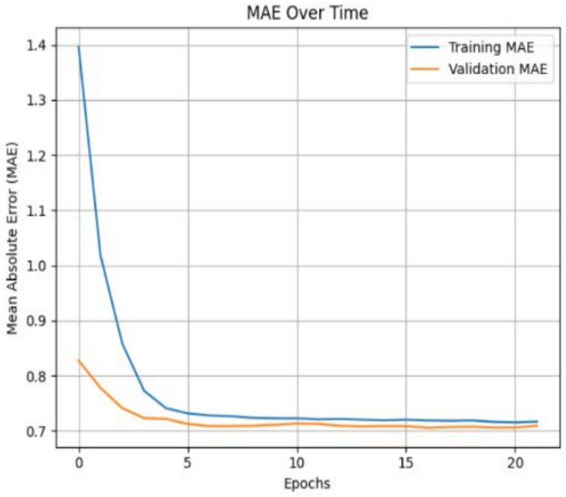


1. **Tag Input Layer:**
   Inputs movie tags (TF-IDF encoded).
2. **Genre Input Layer:**
   Inputs movie genres (encoded).
3. **Concatenate Layer:**
   Merges the outputs of the Tag and Genre input layers.
4. **Dense Layer 1 (32 units, ReLU):**
   Learns feature interactions with 32 neurons.
5. **Dropout Layer 1 (Rate = 0.6):**
   Prevents overfitting by randomly dropping 60% of neurons.
6. **Dense Layer 2 (16 units, ReLU):**
   Refines features with 16 neurons.
7. **Dropout Layer 2 (Rate = 0.6):**
   Another dropout layer to enhance model generalization.
8. **Output Layer (1 unit, Linear):**
   Outputs a single predicted rating or score.

This architecture combines tags and genres to predict user ratings for movies.

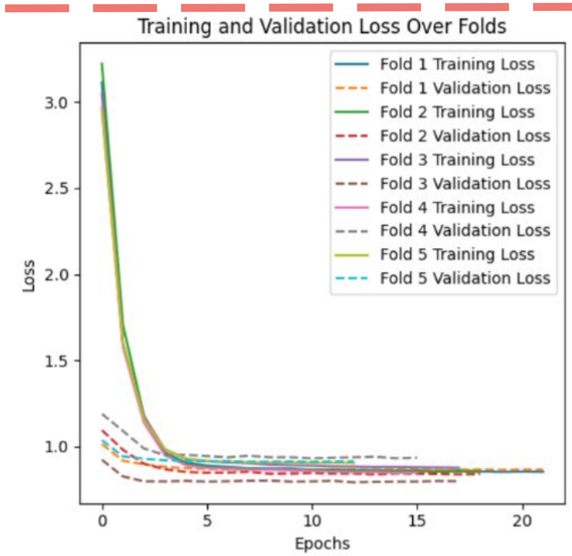
Model Performance & Results

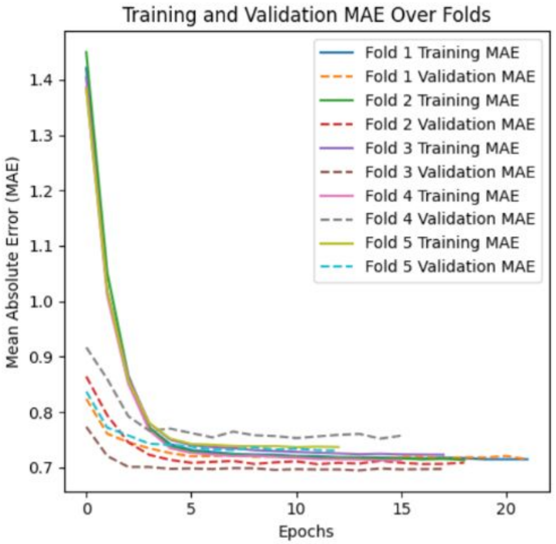




Neural Network final MAE (Mean Absolute Error) is: 0.7058

Cross Validation average MAE is: 0.7213





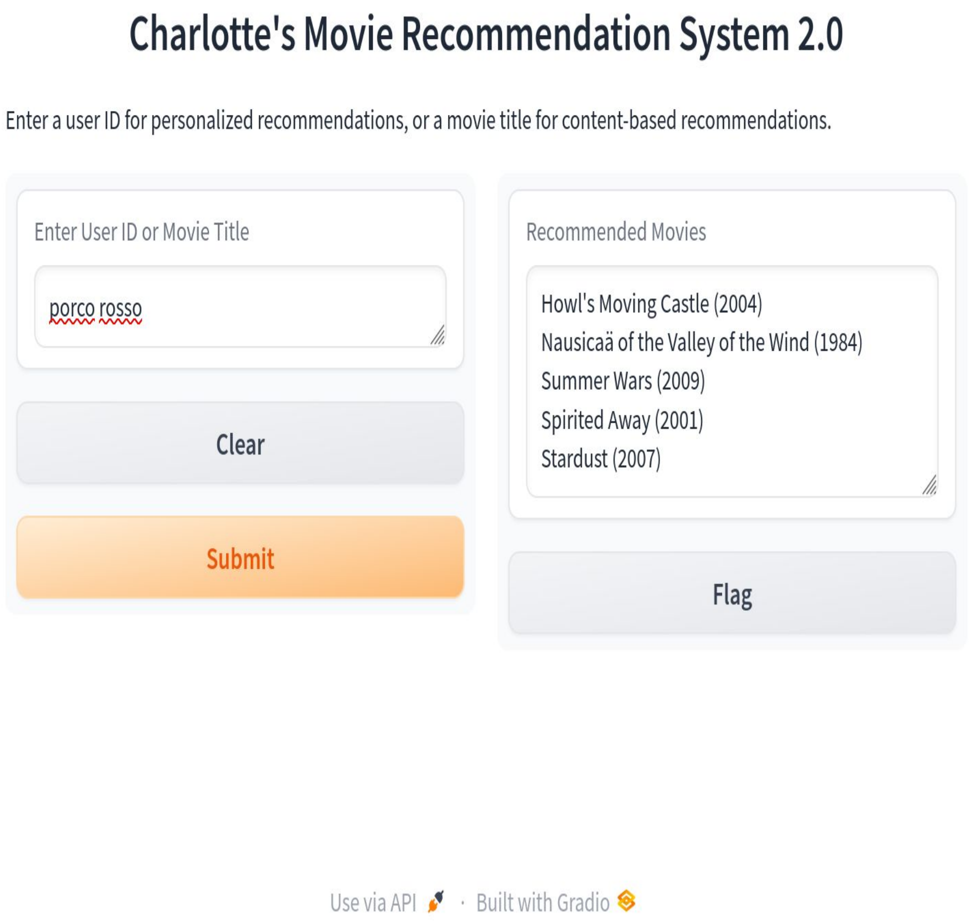
Transformer-Based Recommendation System

Embedding Generation:

Use the SentenceTransformer model to generate **genre and tag embeddings** for each movie, capturing the semantic meaning of the content.

Content-Based Recommendations:

Calculate **cosine similarity** between the movie embeddings to recommend movies based on semantic similarity of genres and tags.



Combined Recommendations:

Mix **genre/tag-based recommendations** with **director-based suggestions** for a more diverse set of recommendations.

Gradio Interface:

Build a **Gradio Interface** where users can input a movie title or user ID to receive personalized or content-based recommendations.

Differences Between System 1.0 & System 2.0

Feature	System 1.0 (Weighted Genre/Tag)	System 2.0 (Transformer-Based)
Recommendation Method	Genre/Tag and Director with Weights	Embeddings (Genre/Tag) and Director
Customization	Allows adjustment of genre/tag weights	Fixed, no weighting needed
Complexity	More complex with weight sliders	Simpler, no weight adjustments
Recommendation Type	Combines genre/tag similarity with director-based suggestions	Embedding-based recommendations for more accurate content matching
Intended Audience	Users who want control over genre/tag weight	Simple, “plug-and-play” for all users
Computational Efficiency	Slightly less efficient due to real-time weight adjustments	Efficient with pre-computed embeddings
Semantic Search	Limited, as weights affect relevance	Strong semantic matching through embeddings

Summary & Recommendations

Improved Personalization: The project developed a movie recommendation system using collaborative filtering, neural networks, and transformer-based models to enhance recommendation accuracy.

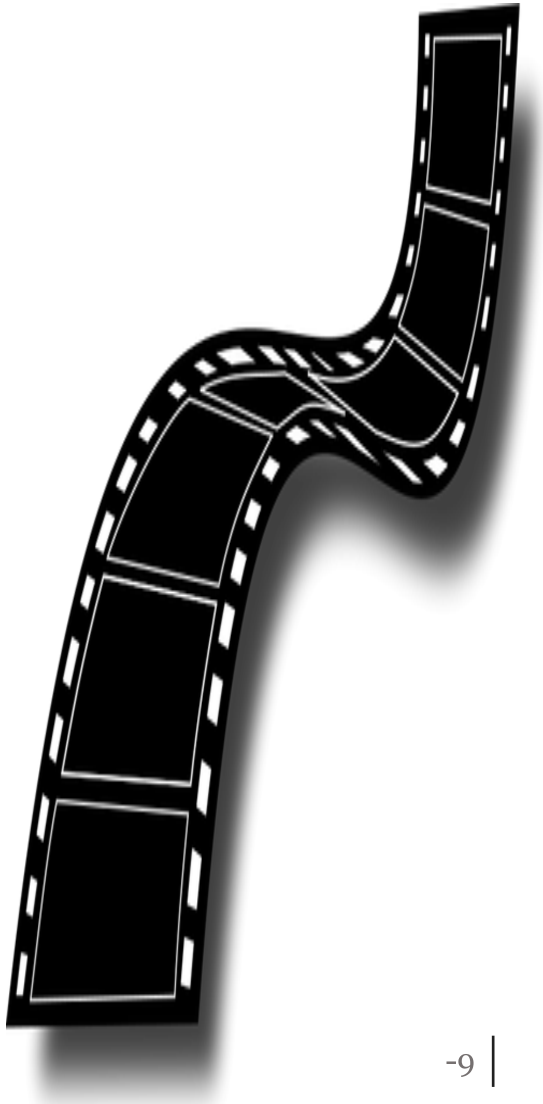
Scalable Approach: By using pre-trained models and advanced machine learning techniques, the system efficiently handles large datasets and provides high-quality recommendations.

Hybrid Recommendation System: The final system integrates both user behavior (SVD) and content-based (Transformer) approaches for more diverse and personalized recommendations.

Increase User Engagement: Offer personalized movie recommendations to enhance user satisfaction and retention.

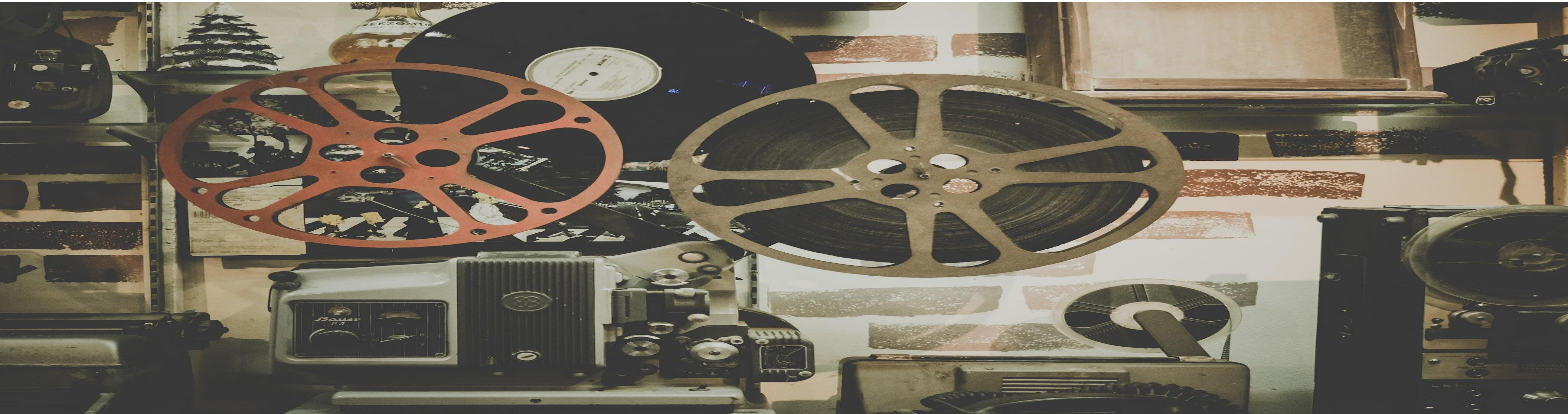
Leverage Data for Insights: Use the recommendation system's data to identify customer preferences and tailor marketing strategies accordingly.

Monetize through Premium Recommendations: Provide exclusive or premium content recommendations, encouraging users to subscribe to higher-tier services.



Next Steps

In the future, the system could be improved by adding real-time user feedback and incorporating multi-modal data such as movie posters and trailers. These updates would help make the recommendations more accurate and personalized.




THANK
YOU !

A little Easter egg has been included at the end of the presentation—a QR code for trying out my movie recommendation system! It can be scanned to interact with the system.

SCAN ME

Version 2.0

MyVVhhvf