Project Proposal: Netflix Recommendation System

Problem Statement
In an era of overwhelming content choices, Netflix users often struggle to find shows and movies that align with their preferences. This leads to user dissatisfaction and potential churn as users may turn to other platforms that better cater to their taste. The goal is to develop a recommendation system that effectively suggests Netflix titles to users, enhancing their viewing experience and satisfaction.

Context
Netflix, as a leading streaming service, offers a vast library of titles, ranging from movies to TV series across various genres. With such a wide array of options, users can find it difficult to discover content that matches their interest. A personalized recommendation system can streamline this process by analyzing user behavior and content attributes to suggest titles that each user is likely to enjoy.

Criteria for Success
1. User Engagement: Increase in the number of titles watched by users.
2. User Satisfaction: Positive user feedback and a decrease in time spent searching for content.
3. Recommendation Accuracy: High relevance of recommendations measured by user interaction with suggested titles.
4. Scalability: The system should efficiently handle Netflix's large and growing content library.

Scope of Solution Space
- Content-Based Filtering: Recommending titles based on similarities in genres, actors, directors, and other attributes from the Netflix dataset.
- Collaborative Filtering: Suggesting titles based on user behavior patterns. Leveraging data such as viewing history and rating.
- Hybrid Approaches: Combining content-based and collaborative filtering methods to improve recommendation accuracy.

Constraints
- Data Limitations: The dataset includes metadata like title, genre, cast, director and release year, but lacks user-specific data such as viewing history or ratings.
- Scalability: The system needs to handle Netflix's extensive and continuously growing catalog efficiently.

Stakeholders
- Netflix Users: Primary beneficiaries who will experience improved content recommendations
- Netflix Management: Interested in user engagement metrics and reducing churn.
- Content Creators: Gain insight into content popularity and viewer preferences.

Data Sources

The primary data source for this project is the 'netflx_titles.csv' file. This dataset includes:
- Title: The name of the show or movie.
- Genre: Categories describing the content.
- Cast: List of main actors.
- Director: The director of the title.
- Release Year: The year the title was released.
- Rating: Audience rating for the title.
- Duration: Length of the show or movie.
- Description: Brief summary of the content.

Data Acquisition
- The 'netflix_titles.csv' data set can be downloaded from public repositories such as kaggle.

Methodology
1. Data Preprocessing: Clean and process the data to handle missing values, inconsistencies, and format the data for analysis.
2. Feature Engineering: Extract and create meaningful features from the data, such as genres, actors and directors.
3. Model Development:
   a. Content Based Filtering: Develop a recommendation model that suggests titles similar to those a user has interacted with, based on content features.
4. Evaluation and Optimization: Use metrics such as precision, and recall to evaluate and fine-tune the models.

Deliverables
1. Code Repository: Well-documented code for the data preprocessing, feature engineering, model development and evaluation.
2. Technical Report: A detailed paper outlines the methodology, findings, model performance and recommendations.

3. Presentation Deck: A slide deck summarizing the project for stakeholders including visualizations and key insights.