**Project Proposal**

**Data source**

In today's digital age, the demand for personalized movie recommendations is on the rise. To address this need, we propose leveraging the extensive MovieLens 25M dataset provided by GroupLens. This dataset comprises 25 million movie ratings and one million tag applications attributed to 62,000 movies and 162,000 users. Additionally, it includes tag genome data with 15 million relevance scores across 1,129 tags.

**Why this source?**

I often find myself in a dilemma when it comes to choosing a movie on Netflix. The process of scrolling through countless options can sometimes take longer than actually watching the movie, and there's always that lingering uncertainty about whether I'll enjoy my choice. That's why the MovieLens 25M dataset resonated with me. With its vast collection of movie ratings and tag applications, it holds the key to creating a personalized movie recommendation system that could better utilize time to actually watch and enjoy the movie. My goal is to develop an algorithm that's not just efficient but also intuitive and insightful, capable of understanding my preferences with just a simple input, like a movie title, and returning 10 similar movies with ratings tailored specifically to my tastes. With MovieLens 25M as my foundation, I'm excited to enhance the movie-watching experience for users like myself.

**Features of the dataset**

I'm planning to combine two datasets: one with movie details like ID, title, and genre, and another with user ratings. By doing this, I can create a personalized movie recommendation system using collaborative filtering. This method looks at what similar users liked to suggest movies I might enjoy. I'll use features like movie ID, title, rating, and genre, along with a "similarity vector" to measure how much users' tastes align. I anticipate generating tailored suggestions that resonate with my movie-watching preferences, based on the viewing habits of users with similar tastes.

**Methodology**

**Data Preprocessing:** Cleanse and preprocess the datasets to handle missing values, outliers, and inconsistencies. Normalize numerical features and encode categorical variables (e.g., genre) for compatibility with recommendation algorithms.

**Data Integration:** Merge the movie details and user ratings datasets based on common identifiers (e.g., movie ID) to create a comprehensive dataset for analysis.

**Feature Engineering:** Extract relevant features from the integrated dataset, including user preferences, movie attributes, and similarity metrics.Generate additional features, such as user-item interaction matrices, to facilitate collaborative filtering.

**Model Selection:** Evaluate various recommendation algorithms, including collaborative filtering (e.g., user-based, item-based), content-based filtering, and hybrid approaches.Choose the most suitable algorithm(s) based on performance metrics, scalability, and interpretability.

**Model Development:** Implement the selected recommendation algorithm(s) using appropriate libraries or frameworks (e.g., scikit-learn, TensorFlow). Train the models using the preprocessed dataset and fine-tune hyperparameters to optimize performance.

**Evaluation:** Assess the performance of the developed recommendation system using evaluation metrics such as precision, recall, and mean average precision (MAP).