# DATA SCIENCE OF PREDICTING IMDB SCORES

## INTRODUCTION:

Certainly, developing a machine learning model to predict IMDb scores for movies is an exciting project. Here's a step-by-step plan to guide you through the process:

---

1. Data Collection:

- Gather a dataset that includes features such as genre, premiere date, runtime, language, and IMDb scores for a diverse set of movies. You can use IMDb's official dataset or explore publicly available movie datasets.

2. Data Exploration and Visualization:

- Conduct exploratory data analysis (EDA) to gain insights into the dataset.
- Visualize data distributions, correlations, and trends using plots and graphs.
- Identify any patterns or relationships between features and IMDb scores.

3. Feature Engineering:

- Create relevant features that could impact IMDb scores:
  - Extract release year from premiere date.
  - Encode categorical variables like genre and language using one-hot encoding or embeddings.
  - Consider creating interaction features or aggregations.

4. Data Splitting:

- Split the dataset into training, validation, and test sets. Common splits are 70-80% for training, 10-15% for validation, and 10-15% for testing.

5. Model Selection:

- Choose machine learning models suitable for regression tasks. Options include:

  - Linear Regression

  - Decision Trees

  - Random Forest

  - Gradient Boosting (e.g., XGBoost, LightGBM)

  - Neural Networks (if you have a large dataset)

6. Model Evaluation:

- Evaluate model performance on the validation dataset using appropriate regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared.

- Select the best-performing model.

7. Fine-tuning:

- Based on validation results, fine-tune your model, possibly by adjusting hyperparameters or adding/removing features.

8. Deployment (Optional):

- Deploy the trained model as a web application, API, or any other suitable platform for users to predict IMDb scores of movies.

9. Presentation and Reporting:

- Create a presentation or report summarizing your findings, methodology, model performance, and insights gained from the analysis.

---

## Design thinking of IMDb movie score prediction project:

1. Empathize:

- Understand the needs and preferences of movie enthusiasts who want better movie recommendations.

- Conduct user interviews or surveys to gather insights into what users value in movie recommendations.

2. Ideate:

- Brainstorm various features and factors that could influence IMDb scores. This aligns with feature engineering.

- Generate ideas for different regression models that could predict IMDb scores effectively.

3. Prototype:

- Create a prototype of the IMDb score prediction system:

  - Develop a data preprocessing pipeline to handle missing values and encode categorical features.

  - Implement a basic regression model (e.g., Linear Regression) as a prototype.

  - Visualize how the model's predictions could be integrated into a movie recommendation system.

4. Test:

- Collect a sample of movie enthusiasts and have them interact with the prototype.

- Gather feedback on whether the predictions align with their movie preferences and if the recommendations are helpful.

5. Test (Again):

- Launch the movie recommendation system to a wider audience.

- Collect feedback and monitor how users interact with the system.

6. Collect Data:

- Continuously gather data on user interactions and movie preferences to refine the recommendation system further.

7. Share and Scale:

- Share the success of the IMDb score prediction and recommendation system with the user community.

- Scale the system to accommodate a larger user base and potentially expand its capabilities.

---

## Conclusion:

- Summarize the key achievements and contributions of your machine learning model in predicting IMDb scores for movies.

---