**IMDB score predicition**

# **IMDB : (**The Internet Movie Database)

The Internet Movie Database (IMDb) is an online database containing information and statistics about movies, TV shows and video games as well as actors, directors and other film industry professionals. This information can include lists of cast and crew members, movie release dates and box office information, plot summaries, trailers, actor and director biographies and other trivia.

Information on IMDb comes from a variety of sources, such as filmmakers, film studios, on-screen credits and other official sources. However, much of the information comes from IMDb users themselves, who can submit facts in a wiki-style format. Unlike traditional wiki sites, IMDb always authenticates information before it appears online -- although errors do show up, and the website allows users to report possible mistakes so they can be fixed.

Users can also submit reviews of movies and TV shows on a one to ten scale, which are then used to create a weighted mean of all user reviews to be displayed on the movie or TV show's page. Also, these ratings are used to compile IMDb's Top Rated and Lowest Rated lists of movies and TV shows.

IMDb was first published in 1990 by Col Needham, a computer programmer, as a group of scripts which allowed users to search a list of film credits that were compiled with the help of a USENET group. The database was incorporated in 1996. In 1998, Internet Movie Database Ltd. became a subsidiary of Amazon, which uses it as a way to advertise movies and DVDs which it sells on its main website.

In 2002, IMDb launched its paid subscription service for film industry professionals, IMDbPro. IMDbPro allows actors, directors and other professionals to manage their IMDb page, upload their résumé and access more extensive information not available on the free version of the website.

In addition to the IMDb website, the company offers a mobile app which is available on iOS, Android, and Amazon Kindle Fire devices. The website has also been translated into German, French, Spanish, Danish, Finnish, Hungarian, Italian, Polish, Portuguese and Romanian.

Features

The title and talent pages of IMDb are accessible to all users, but only registered and logged-in users can submit new material and suggest edits to existing entries. Most of the site's data has been provided by these volunteers. Registered users with a proven track record are able to add and make corrections to cast lists, credits, and some other data points. However, the addition and removal of images, and alterations to titles, cast and crew names, character names, and plot summaries are subject to an approval process; this usually takes between 24 and 72 hours.

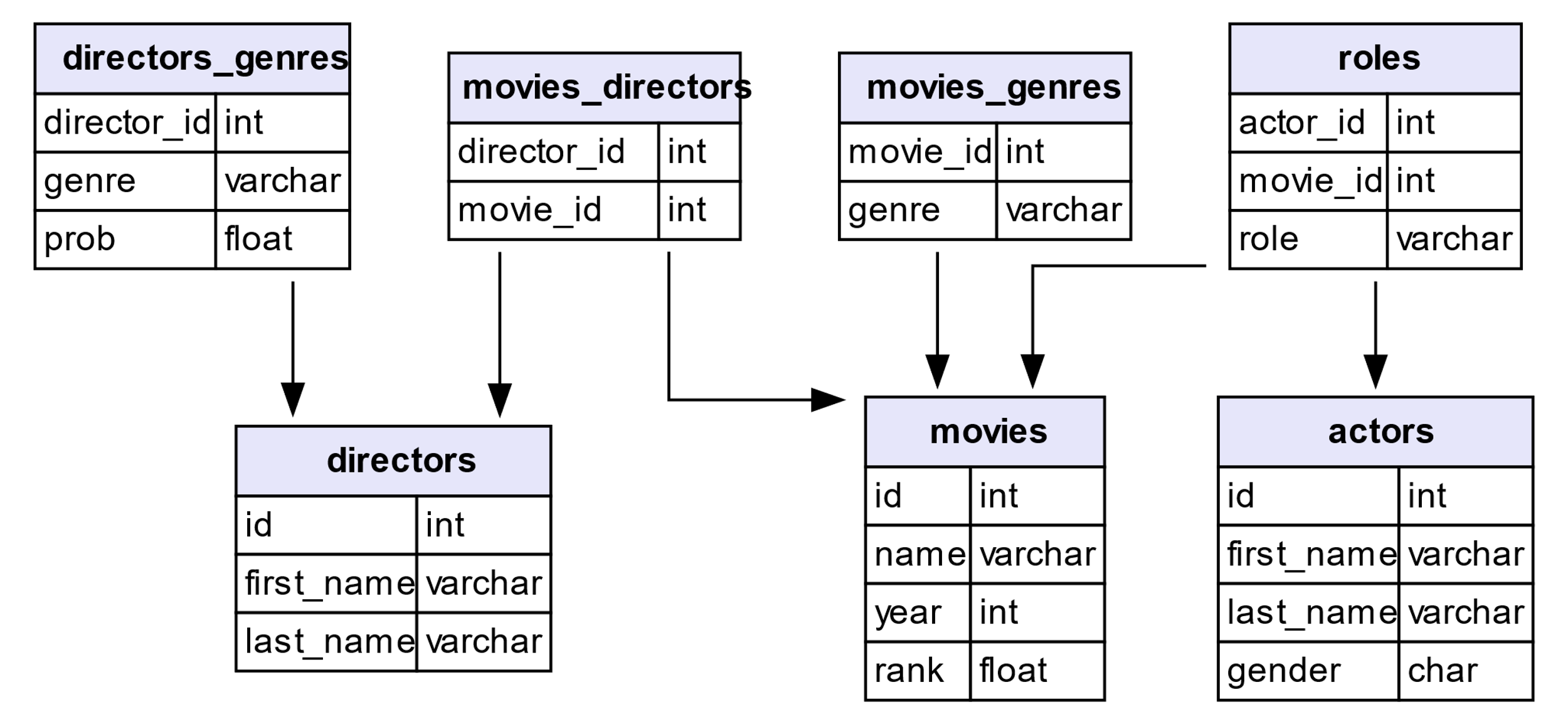
On October 2, 2007,[citation needed] character filmographies were added. Character entries are created from character listings in the main filmography database, and as such do not need any additional verification by IMDb staff. They have already been verified when they are added to the main filmography.

Registered users can choose their username, and most are pseudonymous. There is no single index of contributors, no index on each profile page of the items contributed, and—except for plot synopses and biographies—no identification of contributors to each product's or person's data pages. Users are also invited to rate titles on a scale of 1 to 10, and the totals are converted into a weighted mean-rating, with filters in place to mitigate ballot-stuffing.

User profile pages show a user's registration date and, optionally, their personal ratings of titles. Since 2015, "badges" can be added showing a count of contributions. These badges range from total contributions made to independent categories such as photos, trivia, and biographies. If a registered user or visitor is in the entertainment industry and has an IMDb page, they can add photos through IMDbPRO

**Data Source:**

Obtain a dataset with movie information, including features like genre, premiere date, runtime, language, and IMDb scores. You can find such datasets on platforms like Kaggle or various open data sources.

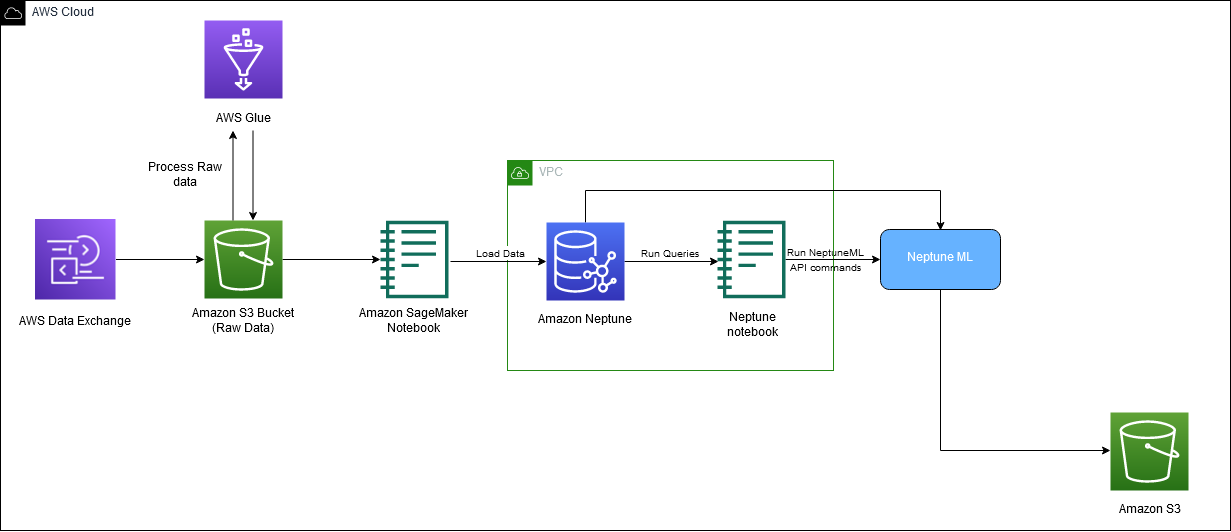


**Data Preprocessing:**

Clean the data by handling missing values, outliers, and duplicates. Convert categorical features like genre and language into numerical representations, such as one-hot encoding or label encoding.

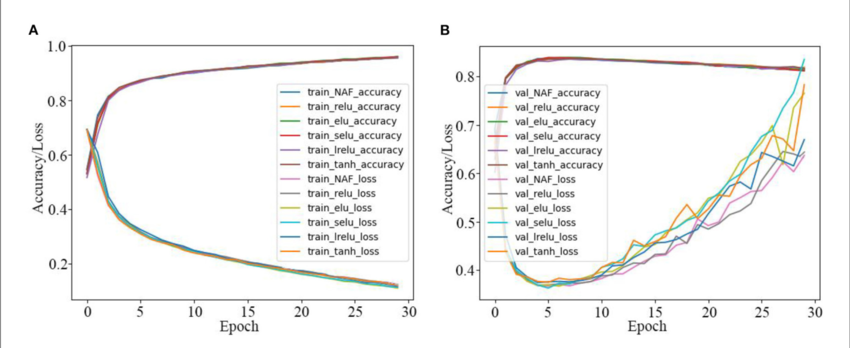
Feature Engineering: Extract relevant features from the data that could potentially influence IMDb scores. For example, you could create new features like the release year, analyze sentiment from movie descriptions, or categorize movies into genres based on keywords.

Model Selection: Choose appropriate regression algorithms for predicting IMDb scores. You mentioned Linear Regression and Random Forest Regressor, which are good choices. You can also consider other regression models like Gradient Boosting, Support Vector Regression, or Neural Networks, depending on the complexity of your data.



**Model Training:**

Split your dataset into training and testing sets, and then train your selected regression model(s) on the training data.



**Evaluation:**

Evaluate the model's performance using regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (coefficient of determination). These metrics will help you assess how well your model is performing in predicting IMDb scores.