INTRODUCTION

The Internet Movie Database(IMDB) is an online database of information related to movies,TV shows,celebrities,genre,reviews,etc.The IMDB website enables registered users to rate different movies,TVshows and actors on a scale of 1 to 10. It also enables users to search different movies or TV shows of different genres on a single platform.

REQUIREMENTS

Prime focus of this project is to provide a generic functional database of IMDB to access entertainment industry websites online.

For this purpose, we have gathered data from IMDB website and have performed different SQL queries keeping in mind the user's perspective and expectations.
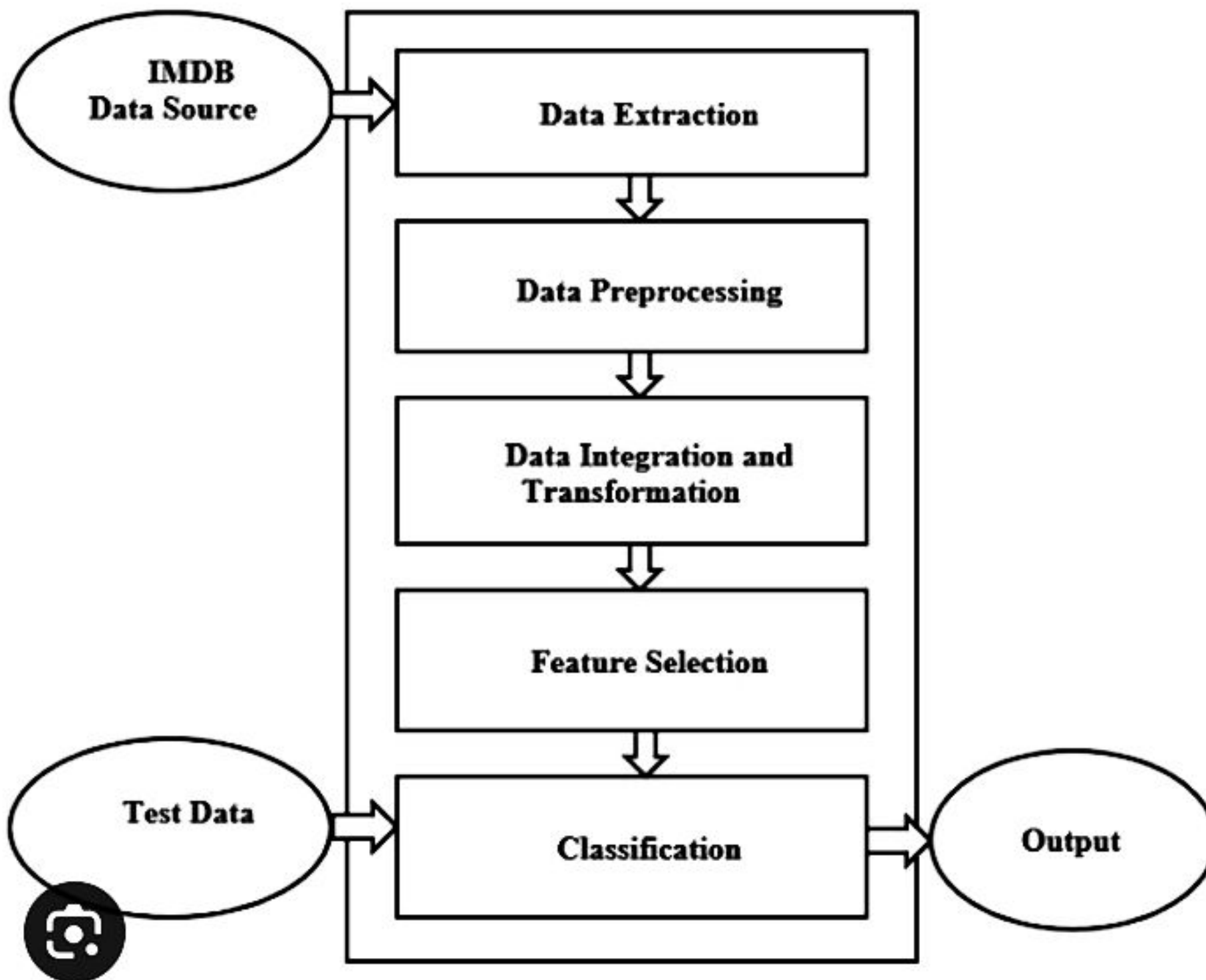
Approach

- Download and Data open in excel
- Find null and duplicate values and clean all data
- Data processing and solved as per asked problems.
- Use filter ,pivot table, sum if, average if, count if and other functions to give answer of asked questions
- Create charts to for easy and meaningful data representation
- Create report in ppt format and submit the project

Problem Definition: The problem is to develop a machine learning model that predicts IMDb scores of movies available on Films based on features like genre, premiere date, runtime, and language. The objective is to create a model that accurately estimates the popularity of movies, helping users discover highly rated films that match their preferences. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

```
IMDB
Data Source  →  Data Extraction
                      ↓
                 Data Preprocessing
                      ↓
                 Data Integration and
                 Transformation
                      ↓
                 Feature Selection
                      ↓
Test Data  →  Classification  →  Output
```

Design Thinking:

1. Data Source: Utilize a dataset containing information about movies, including features like genre, premiere date, runtime, language, and IMDb scores.

2. Data Preprocessing: Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.

3. Feature Engineering: Extract relevant features from the available data that could contribute to predicting IMDb scores.

4. Model Selection: Choose appropriate regression algorithms (e.g., Linear Regression, Random Forest Regressor) for predicting IMDb scores.

5. Model Training: Train the selected model using the preprocessed data.

6. Evaluation: Evaluate the model's performance using regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.