1. Name of the project and team members

Movie Rating Prediction and Influential Factors Analysis Based on TMDB Data

Rui（isabella） wang & I-Chen（Iris）Li

2.What problem are you trying to solve?

This project aims to analyze the key factors that influence movie ratings and to develop a model capable of predicting movie ratings. By studying the relationships between features such as movie genre, actors, directors, release time, and popularity with ratings, it seeks to help audiences better understand the patterns of movie ratings and provide a reference for predicting the ratings of new films.

3.How will you collect data and from where?

To achieve this goal, relevant movie data was crawled through the TMDB (The Movie Database) API, covering basic movie information, including title, release date, genre, runtime, and language; cast and crew information, such as directors and main actors; rating information, including average rating and number of votes; as well as movie popularity and trending metrics. During the data collection process, Python's web libraries were used to fetch the API data, and the retrieved JSON data was cleaned and stored using relevant libraries for subsequent analysis.

4.What analysis will you do and what visualizations will you create?

In the data analysis section, descriptive statistical analysis will be conducted first to understand the distribution of ratings, popular genres, and trends in ratings over time. Additionally, correlation analysis will be performed to explore the relationships between movie features and ratings. Based on these analyses, regression models such as linear regression and random forests will be used to predict movie ratings. For visualization, we plan to generate histograms and box plots of rating distributions, bar charts of genre popularity, scatter plots showing the relationships between features and ratings, and feature importance plots from the regression models, thereby presenting the analysis results and model insights in an intuitive manner.