Project background and goals

- In the current entertainment industry, the criteria for film success are increasingly dependent on audience ratings and feedback. Our goal is therefore to understand and predict the relationship between a film's ratings success and its producers and distributors. By analyzing the TMDB all-movie dataset, we aim to identify and predict which factors most determine a movie's rating success. We will achieve this through a two-stage machine learning model: first, using a linear regression model to explore the relationship between movie characteristics and its rating success; second, using a deep learning regression model for deeper analysis and prediction.

Week 1: Data Processing

- Confirm project goals and division of labor.
- Collect all movie data sets in TMDB and conduct preliminary data exploration.
- Data preprocessing: including missing value processing, outlier detection and correction, and feature extraction.

Week 2: Development and Analysis of Linear Regression Models

- Based on the selected features, develop a linear regression model to analyze the relationship between each feature and movie rating success.
- Use cross-validation to evaluate the performance of linear regression models and determine the best combination of features.
- Based on the results of the linear regression model, filter out the features most important for successfully predicting movie ratings.

Week 3: Development and Analysis of Deep Learning Models

- Design and implement a deep learning regression model integrating the best predictive features selected in the second week.
- Train a deep learning model on a subset of the dataset and evaluate it on an independent test set to verify the model's predictive capabilities.
- Adjust model parameters and structure to improve prediction accuracy and model generalization ability.

Week 4: Project Summary

- Comparatively analyze the performance of linear regression models and deep learning models.
- Summarize project findings and prepare final report. The report will detail the project's methodology, main findings, model comparisons, and prospects for future work.
- Create project presentation materials, including key findings and conclusions, and how our findings can be used to predict a film's critical success.

Expected results

- By analyzing and comparing different machine learning models, we expect to gain deep insights into which movie features most influence its rating success.
- Based on our models and analysis, we plan to develop a tool capable of predicting a film's potential critical success before its release.